

CONTENTS

What is the Congressional Record?	1
Why is there a daily and a bound edition of the Record?	1
Is that all there is to understanding the Record?	2
Why is the Record so hard to use?	4
How does ProQuest meet these challenges?	4
Our metadata philosophy addresses the size	4
How is this possible?	5
Metadata matters to relevant Search results	5
CIS indexing harmonizes subjects across the years	6
The citation challenge	6
Conclusion	6

What is the Congressional Record?

Marc Anderson, Andrea Severson, and Diane Smith

In the American imagination there are few more iconic images of Congress than Jimmy Stewart's desperate filibuster in Frank Capra's classic film *Mr. Smith Goes to Washington*. For today's researcher it is difficult to balance that image with the picture of a lone individual talking to an essentially empty chamber as televised on C-SPAN. What is the truth? What really happens on the floor of Congress?

The closest that anyone can get to the answer to this question (other than serving as a legislator) is by reading the *Congressional Record*. The *Record* is the public account of all the legislative debate and wrangling that happens every day in Washington. As the official record of the proceedings, debates, and activities of Congress, the *Record* is a substantially verbatim account of all discussion occurring on the floor of Congress. The *Record* also contains inserted materials, communications from the president and the executive branch agencies, memorials, petitions, and information and materials on legislation being introduced or passed.

Why is there a daily and a bound edition of the Record?

One of the first discoveries for the researcher using the *Record* is that there are two editions -- a daily edition and a bound, permanent one -- both of which are printed by the Government Printing Office (GPO). The question of "Which one to use?" and "Why are there two versions?" naturally arises in the user's mind.

The daily edition reports each day's proceedings in Congress and is printed by the following morning. In many ways the *Record* is the newspaper of record for the Hill -- an analogy that is enhanced by the fact that it is printed on newsprint. Currently, the daily edition is divided into four distinct and separately paginated parts: one covering House proceedings; one covering Senate proceedings; one entitled "Extension of Remarks," and one called the "Daily Digest" (not to be confused with the daily edition). The "Extension of Remarks" section includes additional materials inserted into the public record by legislators. The "Daily Digest" is a brief overview of the key activities in both the House and Senate for the purpose of quick perusal. Every other week GPO produces a paper index to the content of the last two weeks' daily editions.

At some time after the end of each annual session of Congress, all of the daily editions are compiled, re-paginated, and printed as a permanent, bound edition. These editions usually take several years to reach publication. (For example, the bound volume for 2003 is just being published in 2008 and the most recent index volume covers the proceedings of 1997.) The bound edition is made up of one volume per session of Congress and each volume is published in multiple parts.

The bound edition differs from the daily edition in significant ways:

- Pagination is continuous.
- Text is edited, revised, and rearranged (Members of Congress have the authority to edit material after the text was originally published in the daily edition).
- The prefixes of H, S, and E before the daily page numbers are dropped.
- High-quality paper stock is used.

Is that all there is to understanding the Record?

No, of course not!

The government has published the *Record* for more than 130 years and as technologies and policies have evolved, so has the *Record*. There have been changes in layout, content, indexing practices, and in the formats sent to libraries (paper, fiche, electronic, and CD-ROM). Fortunately there are many excellent written descriptions of the various idiosyncrasies that one encounters when using the *Record*. The annotated bibliography below suggests further readings on the topic. And figures 1 and 2 provide a chronology of some of the format and content changes in congressional publishing practices over the years.

Selected Annotated Bibliography

McKinney, Richard. "An Overview of the *Congressional Record* and Its Predecessor Publications." *LLSDC's Legislative Source Book*. Law Librarians' Society of Washington, D.C. Last revised in September, 2006. www.llsdc.org/sourcebook/cong-record.htm

Most people looking for background on the congressional debates will need read no further than this article which details each of the publications, the Daily Digest, and Appendix. Though there are a few omissions (for example, there is only the briefest mention of inserted remarks) this article has the benefit of being easily accessible and revised periodically.

McPherson, Elizabeth Gregory. "Reporting the Debates of Congress." A paper prepared for the Twenty-sixth Annual Convention of the National Association of Teachers, Detroit Michigan, Dec. 30, 1941. Appears in the *Quarterly Journal of Speech*, v. 28, April 1942. p. 141-148. [Reprinted in the *Congressional Record*, v. 88, June 10, 1942. A2182-A2185]. See also Mundt, Karl. Reporting the Debates of Congress. Remarks in the Senate. *Congressional Record*, v. 88, June 10, 1942. p. A2182-A2185.

A lively article with wonderful descriptions of early debates (1789-1873) and how they were recorded, or not. McPherson gives the reader the feel for the politics involved with whether or not certain parties or individuals were accurately recorded. For those interested in more background on the Senate history of recording debates, see McPherson, Elizabeth G. "The Southern States and the Reporting of Senate Debates, 1789-1802." *Journal of Southern History*. Vol. 12, No. 2, (May, 1946), pp. 223-246.

U.S. Congressional Research Service. *Congressional Record: Its Production, Distribution, and Accessibility* (98-266 GOV; CRS-2006-GVF-0754; Oct. 27, 2006), by Mildred Amer. 2p.

A very brief overview of the present day Record. Includes description of all of the parts of the record, as well as amendments, additional statements (Senate) and the Extension of Remarks (House).

U.S. Congressional Research Service. *Congressional Record: Content, History and Issues* (93-60 GOV; CRS-1993-GOV-0022; Jan. 14, 1993), by Mildred Amer. 24p.

A good, relatively brief, overview of the *Congressional Record*. This report has all of the basic material covered, plus a section on distinguishing spoken and unspoken remarks (a.k.a. use of the bullet).

U.S. Congressional Research Service. *Congressional Record* (86-152 GOV; CRS-1986-GOV-0011; Oct. 21, 1986) by Mildred Amer. 78p.

For those interested in details about the *Congressional Record*, this is the mother lode. While the first thirty pages are essentially replicated in the 1993 report cited above, Appendix D gives a guide to the *Annals of Congress*, the *Register of Debates*, the *Congressional Globe*, and the *Congressional Record* by detailing, volume by volume, not only which volumes cover which year/session/congress, but also the "Commentary Notes" give detail as to when and how it was published and whether or note additional material is included. For example, page 57-8 notes that the "Register of Debates ... together with an appendix, containing the most important state papers and public documents to which the session has given birth: to which are added the laws enacted during the session, with a copious index to the whole."

What is the Congressional Record?

figure 1

Chronology of Events

- 1794** Senate opens to spectators on a trial bases.
- 1795** (December) Senate opens to spectators.
- 1833** *Congressional Globe* begins publication and is the first journal to list each step in every legislative measure
- 1834** Gales and Seaton publish the *Annals of Congress*, covering 1789-1824.
- 1837** *Register of Debates* ends.
- 1851** *Congressional Globe* becomes essentially verbatim due to improvements in note-taking.
- 1855** *Congressional Globe* reporters get paid at public expense.
- 1873** Congress votes to have debates published by the congressional printer at the Government Printing Office (17 Stat 510).
- 1873** *Congressional Globe* ends publication; *Congressional Record* begins publication; Congress employs "Official Reporters of Debate."
- 1880-1970** Subscription price for the *Congressional Record* set at \$1.50 per month.
- 1920** Congress considers barring printing of inserted speeches, but concludes it is "preferable to let them be printed rather than be compelled to listen to them."
- 1936** The inclusion of illustrations is prohibited without the approval of the Joint Committee on Printing (44 USC 904)
- 1937-54** The Appendix included as separate section (often as separate volumes of the bound edition), containing remarks not actually delivered on the floor and revisions of remarks.
- 1941** *Congressional Record* changes from two columns to three.
- 1947** Daily Digest begins publication.
- 1954-66** Appendix omitted from the bound *Congressional Record*, while continuing to appear in the daily version. However, Appendix materials judged to be germane are still inserted into bound *Congressional Record*.
- 1968** The Extension of Remarks section replaces the Appendix, appearing immediately after the proceedings of the relevant day.
- 1970** Public Printer is authorized to set subscription price for *Congressional Record* based on printing and distribution costs.
- 1979** House begins televising proceedings.
- 1985** House begins using different font to indicate inserted material.
- 1986** Senate begins televising proceedings.
- Source:** <http://www.lexisnexis.com/congcomp/getdoc?CRDC-ID=CRS-1993-GOV-0022>

figure 2

A timeline of changes to the Appendix of the *Congressional Record*

- 1873-1919** (43rd-65th Cong.)
- Appendix appears in separate section after all proceedings text for a session.
 - Appendix section has its own consecutive pagination, independent of proceedings. There is no notation to distinguish from proceedings pagination – all are simple numerals.
 - Appendix has own index
- 1919-1922** (66th Cong., 2nd sess. - 67th Cong.)
- Appendix appears in separate section after all proceedings text for a session.
 - Appendix loses independent pagination, and follows consecutive pagination of proceedings.
 - Appendix no longer has own index.
- 1923-1936** (67th Cong., 3rd sess. – 74th Cong.)
- Appendix no longer appears in the bound *Record* at all.
- 1937-1953** (75th Cong., - 83rd Cong., 1st sess.)
- Appendix reappears in bound *Record*.
 - Appears in one or more volumes following proceedings.
 - Has own independent pagination. In 75th and 76th Congresses, Appendix pages have plain numerals. Starting with 77th Congress, page numerals in Appendix begin with letter "A" (A15, etc.).
 - Appendix is indexed in regular indexes. In 75th and 76th Congresses, index cites to Appendix pages say "Appendix" before page number. Starting with 77th Congress, index cites to Appendix pages just use the "A" page number.
- 1954-1966** (83rd Cong., 2nd sess. – 89th Cong.)
- Full Appendix does not appear in bound edition (continues to appear in daily edition).
 - Items from appendix that are considered "germane" are included in the bound edition, in an Extension of Remarks section following the proceedings and included in the same pagination.
 - The Extension of Remarks items are indexed in regular indexes. "Non-germane" items appearing in the daily Appendix are still indexed (with "A" numbers in citations), though they do not appear in the bound version text. This has apparently led to some libraries keeping the daily version appendices with the bound volumes as a stopgap. However, since "non-germane" material is not truly part of the bound version, we are not including it in our product.
- 1967** (90th Cong., 1st sess.)
- Appendix with separate pagination reappears in bound edition.
- 1968-present** (90th Cong., 2nd sess. – present)
- Extension of Remarks is placed after proceedings for each day.
 - Pagination is continuous from section to section and day to day.
 - Extension of Remarks is indexed along with Proceedings.

Why is the *Record* so hard to use?

From a researcher's perspective there are three main challenges in using the bound and the daily editions. First, the mammoth size of the bound volumes is enough to stop all but the most resolute of researchers. Today the bound *Record* consists of approximately 20 to 30 volumes of proceedings, one or two index volumes, and one "Daily Digest" volume per session -- approximately 30,000 pages of material annually! Usually when researching a topic or a piece of legislation, one must look at many years of congressional activity to be comprehensive. Even with the detailed index, navigating through the *Record* involves retrieving and flipping through a series of very large hardbound books (or microfiche depending on the library's depository status) and the process can quickly become cumbersome and confusing. The total size of the bound *Record* (1873-1997) is approximately 2.5 million pages in more than 1,900 volumes. The predecessor volume -- the *Annals of Congress*, the *Register of Debates*, and the *Congressional Globe* -- add another 200,000 pages with potential research value, depending on the researcher's topic.

The second challenge is the nature of the indexing. Indexes for the bound volumes produced in the past 30 years are extremely detailed with descriptive notations for each entry, but earlier indexes are much less detailed and useful. Also, the indexing vocabulary used varies over time; a single topic may be covered by a number of synonyms or variant terms. A researcher, therefore, continually has to think in synonyms when searching for topics over time or trying to compare the *Record* from different time periods. For example, information that today is indexed under the descriptor "National Security" might be found under descriptors such as "National Defense," some variation of "Military," or even "Militia" in previous years -- different ways of describing the same thing.

Another issue is that the index is constructed at a rather broad level. For example, materials about education in the modern *Record* are typically filed in one large lump under the descriptor "Education." A researcher looking for a more specific education topic ("Head Start," "school vouchers," or "special education," for example) would need to sift through up to a thousand or more entries to find the information they need.

Finally, there is the challenge of citations. Frequently researchers have a daily edition citation and must find the same content in the bound volume. Because of the repagination process previously discussed, a citation in the daily *Record* will not work in the bound *Record*, nor does the reverse work.

How does ProQuest meet these challenges?

Our metadata philosophy addresses the size

Our approach to digitizing the *Congressional Record* is to create a separate bibliographic record for each day that includes all indexing (both traditional CIS indexing as well as the indexing found in the bound *Congressional Record* index volumes -- see figure 3). This allows users to search across as many years as they wish with one search, and locate every day in which their topic was discussed. And researchers may limit their searches to specific content types found within the *Record* (for example, speeches, article and editorials, bill text, tributes). Users will also be able to search for statements made by members of Congress, for specific bills and resolutions, and roll call votes. This will allow users to pinpoint specific topics and types of materials and gather all citations at once, rather than going through volume by volume.

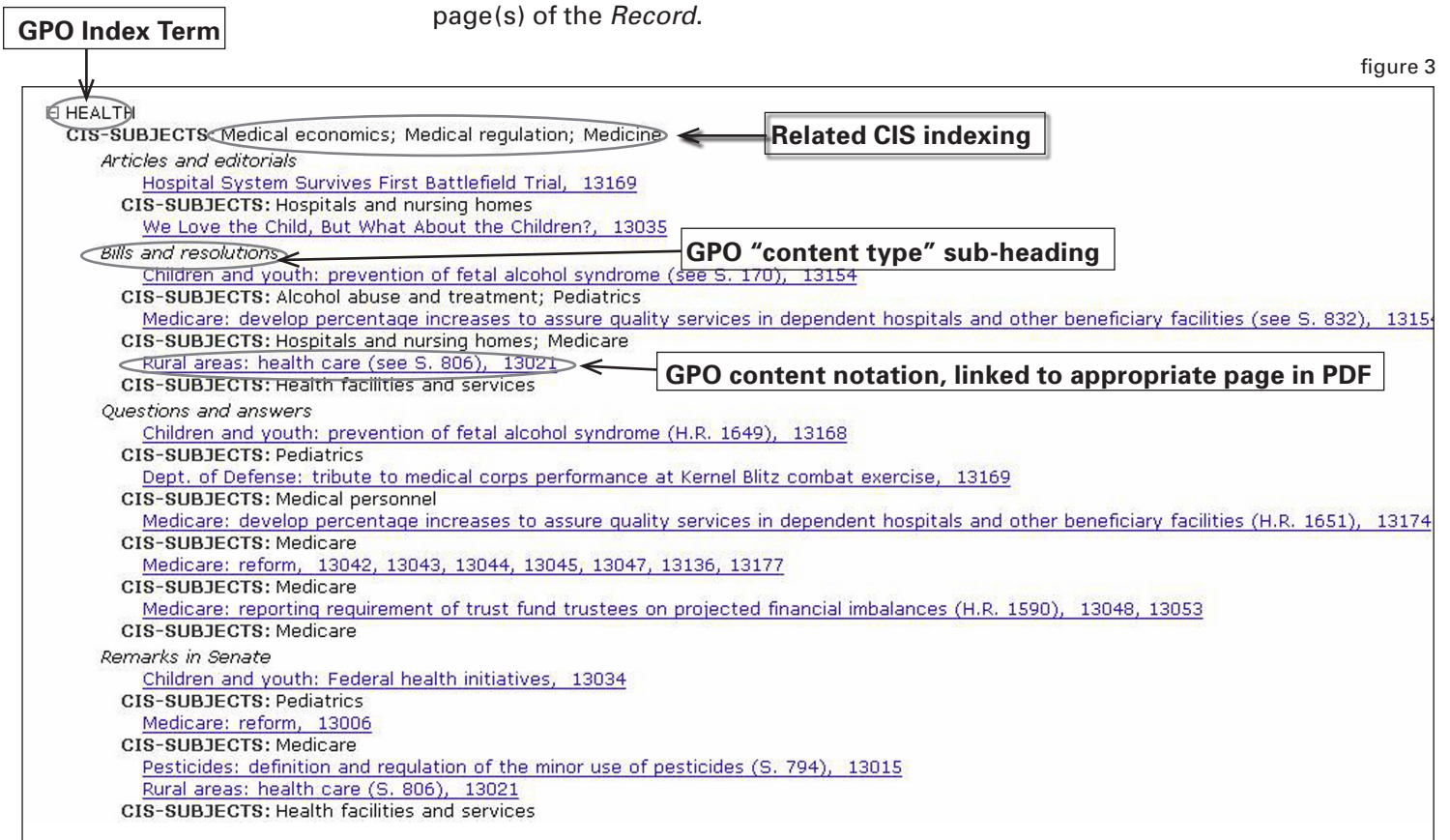
What is the Congressional Record?

How is this possible?

We are doing this by associating the indexing with the page ranges for that day. Let's assume that a page of the bound *Congressional Record* index volume reads as follows:

Corn	12789, 22008
Cotton	456
Soybeans	1378
Wheat	23

As part of our digitization efforts we are dividing every bound volume into days and then capturing the page ranges covered on that day (for example, January 5th covers pp. 998-1475). This allows us to insert all the index terms found in the bound *Congressional Record* index volume into our metadata record for the day (in our example we know that the soybean discussion occurred on January 5th because its pagination falls in the page ranges covered in the bound volumes). Each indexing entry will also include a hyperlink leading directly to the cited PDF page(s) of the *Record*.



Metadata matters to relevant search results

While full-text searching will be available, our approach means that the researcher need not be dependent only on full-text searching, which can lead to many false hits in a publication as wordy as the *Record*. For instance, imagine searching the topic "Vietnam" with a full-text search. The results would list each and every time a legislator discussed Vietnam, as well as every time the Vietnamese conflict was mentioned in any debate on military or foreign policy efforts. Or consider what happens if one does a full-text search on Al Gore's name; one would get hits for every mention of Gore, whether or not he is a speaker or a topic. Using the ProQuest metadata, a researcher will be able to use precise indexing to target search results over as many years as necessary in one search, always retrieving content in which the search topic is a main theme rather than just a passing mention.

CIS indexing harmonizes subjects across the years

In addition, we are adding our CIS indexing to each record to harmonize the GPO terms with our controlled vocabulary – thus allowing across-content searching among the ProQuest digital hearings, documents, reports, prints, and CRS reports. In our metadata records for each day, CIS index terms will be placed alongside the GPO terms and the content notations that they match, weaving the proven and focused CIS vocabulary together with the detailed GPO content notations.

The citation challenge

A problem that many researchers face is that they only have a daily citation and need to translate that to the bound volume. The congressional practice of shuffling content and repaginating the bound volume makes the translation from a daily edition citation to a bound volume very difficult. This research problem is exacerbated for legal researchers because the *Bluebook: A Uniform System of Citation* (18th edition) requires citation to the bound version over the daily, if possible (see section 13.5 of the *Bluebook*).

Our solution to this problem is to develop a translation table between the two editions, using the date as the common element. As mentioned earlier, we are capturing the page ranges as part of the digitization of the bound volumes. With that information we will be able to build an automated look-up table to locate the pages contained in each daily edition. Our system will link a daily citation to its corresponding bound citation, allowing a user to retrieve the bound edition record and the PDF for the pages on which that day falls. The user will then be able to either peruse the day's full text or use the Adobe search capability to search within the day's proceedings for the exact text.

Conclusion

All current *ProQuest Congressional* customers already have the last 22-plus years of the daily edition of the *Congressional Record* as a full text searchable file within their basic subscription. Because the bound edition is considered the official version of the *Record*, we are focusing our efforts on significantly improving access to those volumes. Our digital version allows users to search the *Record* simultaneously with other congressional publications (for example, hearings, reports, and documents) while speeding up and improving their research results by addressing the three problems that researchers have in using the *Record*: its size and complexity, its indexing variability; and its pagination variations.