What is the Serial Set?

One of the more intimidating tasks facing new documents librarians is figuring out what is in those 14,000+ volumes sitting there on the shelf. It reminds me of a Saturday Night Live routine we quoted in college. “What the____ is that thing?” the actors exclaimed as they ‘viewed’ the item in question.

It’s sheer size makes the Serial Set intimidating, and the bindings can make it look more like a tabula rasa than what it really is – one of the most important collections for the study of American history, public policy, women’s issues, and more.

And where else could you find answers to questions like "what language did the Yattassee tribe speak?" (ASP 07 Ind.aff. 113/3); or "what were the names of all the warships that took part in the Jamestown Ter-Centennial Exposition in 1907?" (5403, 60 S.Doc. 735); and do you have a copy of an essay entitled "Anti-Communist Education, Vital Prerequisite To Save the Free World" (12,444, 87 S.Doc. 93).

So how do you get an idea of what is in that thing (or “thang” as they pronounced it on Saturday Night Live)?

When I was starting out at Berkeley,
all I had were the various documents

textbooks that gave rather dry definitions,

but I managed to stumble along – somehow. I would have been much more

informed, and able to really exploit the

resources available in the Serial Set had

some of today’s background material been

available to me then.

**BACKGROUND READING**

Two good resources include: Virginia Saunders’ web document “U.S.

Congressional Serial Set: What It Is and

Its History” ([www.access.gpo.gov/

su_docs/fdlp/history/sset/index.html](http://www.access.gpo.gov/su_docs/fdlp/history/sset/index.html)) that

explains both the contents as well as the physical history of the set; and the

Congressional Information Service (CIS)

“The U.S. Serial Set A Basic Introduction”

(in the Help Text of the print volumes, and also offered through the online help).

While these are both sources that explain a lot of the history of the Serial Set itself, the more engaging material, material that talks about the contents of the Serial Set itself, comes from some of the following sources. It can be a lot of reading, but the potential use for the materials inside the Serial Set comes alive in these readings.

Sources like Suzanne deLong’s, “What Is in the United States Serial Set?” (*Journal of Government Information*, v.23:2 (1996) pp 123-35) which compares the output of four Congresses and gives the reader a flavor of what was going on at each time period by the list of titles included; or a Wooster College website devoted to the various SuDocs classes in the Serial Set ([www.wooster.edu/library/gov/serialset/Introduction.htm](http://www.wooster.edu/library/gov/serialset/Introduction.htm)), which also serves to give details on the serials therein.

There are also articles that explore specific elements of the Serial Set including Charles Seavey’s “Government Graphics: The Development of Illustration in U.S. Federal Publications, 1817-1861” (*Government Publications Review*, v.17 (1990) pp 121-42) in which he reviews the illustrations, discusses their origins, and outlines which agencies took the lead in developing illustrated publications; and Rebecca Wondriska’s “Women and the American Dream, 1900-1925” (*Government Publications Review*, v. 17 (1990) pp 143-57), an annotated bibliography of resources by and about women in the early part of the twentieth century.

There are discussions of the rare items within the Serial Set such as “Documents as Rare Books” (*DttP*, v.14 #4, December 1988, pp 174-79) which reviews not just rare books, but also the ever-popular reports of American explorations. One such report, the Fremont survey, warranted its own article by Stephen Craig Weiss, “The John C. Fremont 1842, 1843-'44 Report and Map” (*Journal of Government Information*, v.26:3 (1999) pp 297-313). The abstract explains the importance of the Report:

“Fremont’s Report was more than a travelers’ guide, it was a U.S.
government publication that achieved
the expansionist objectives of a nation, and provided scientific and economic information concerning the potential of the trans-Mississippi West for pioneer settlement.”

In addition to the material referenced in the articles above, there are reports from the House Committee on Un-American Activities, and gems like the *Official Records of the Union and Confederate Navies in the War of the Rebellion*, a document containing an engraving by a J. A. Whistler, a patent registration from A. Lincoln, the popular *Statistical Abstract of the United States*, *Foreign Relations of the United States*, many of the *Bulletins* of the Bureau of Labor Statistics, and such staples of government as the annual reports of agencies (or secretaries of agencies).

On the more prosaic side, there are many, many – about 240,000 – House and Senate reports, of which more than 100,000 are related to bills that were before Congress. Also included are the *State of the Union* messages, the *Journals* of the House and Senate whose import is documented by Walter Stubbs in “Finding Congressional Journals in the Serial Set” (*Journal of Government Information*, v.24:1 (1997) pp 39-45), and several hundred hearings including those held on the sinking of the Titanic, Bleeding Kansas, the activities of the Ku Klux Klan after the Civil War, and women’s suffrage.

**How Did Publications Get Into the Serial Set?**

Since Congress took responsibility in selecting the items for the *Serial Set*, it’s a fascinating, though somewhat uneven, collection. Congress selected items for several reasons: to provide Members and their staffs with information needed in carrying out their constitutional responsibilities; to see that a permanent record is kept of congressional activity for the use of all interested parties both within, and outside of the Capitol; and to influence public opinion in matters that interest Congress.¹

In his article “Clio’s Blindsport” (*Government Publications Review*, v.13 (1986) pp 67-78), Steven Zink states that:

“For the most part, the Serial Set contains the published working papers of Congress specifically directed to be printed for its use. In the 19th century and the very early 20th century, a considerable amount of executive branch material was included in the Serial Set. Thus, one can find annual reports for such executive agencies as the Bureau of Indian Affairs within the Serial Set during the late 19th century, but will not find such an annual report in the Serial Set of the 1980s.”

Suzanne deLong quotes the CIS® Serial
Set Index when she reports that the Serial Set, taken as a whole, contains:

- Congressional journals, and administrative reports, directories, manuals, and related internal publications;

- Congressional reports on public and private legislation considered during each Congress;

- Reports resulting from congressionally commissioned or conducted investigations;

- Annually submitted reports from Federal executive agencies, reviewing current problems and activities under agency purview;

- Extended series of survey, research and statistical publications developed by executive agencies; and

- Selected annual or special reports of nongovernmental agencies.²

So while there may not be a consistent rationale for what’s in the Serial Set, it all points to Congress and what they needed for their work.

Who Are the Most Likely Users?

I’ve used the Serial Set for most of my career as a documents librarian, and its value is immense to researchers of legislative intent of Congress; American political thought; social, economic and military history; international relations; women’s studies; civil rights; geography; geology; and more. While publications and products in the library world are often geared to a specific educational group (the graduate student, the undergraduate and so on) or to a specific field, the broad swath of subjects covered in the Serial Set makes it an indispensable source for those studying the humanities, social sciences, and many areas in the sciences.

Let’s look at the discipline of science, since most librarians and users would not consider the Serial Set as a relevant collection for such research. Viewing the subject hierarchy for geology you find links to the broader terms earth sciences, physical sciences, and paleontology, and the related terms (“see also” references) earthquakes, glaciers, mines and mineral resources and volcanoes. A subject search for “salmon” (1850-1900) will bring back results ranging from the Fish Commission Bulletin, v. 18, 1898 (3832, 55 H.Doc. 308), Draft of Bill for Protection of Salmon Fisheries in Alaska (3692, 55 H.Doc. 401), Providing Better Protection for Alaska Fur Seals and Salmon Fisheries (2619, 50 S.Rpt. 2687), to the Annual Report of Governor of Alaska, 1899 (3918, 56 H.Doc. 5/21), to name a few.

My own major of linguistics comes up a more than a few times: Bilingual Education Act (12753-6, 90 H.Rpt. 915), The Indians of the Southeastern United
States (10744, 77 H.Doc. 767), and Bulletin 127, Linguistic Material from Tribes of Southern Texas and Northeastern Mexico (10424, 76 H.Doc. 487); the Annual Report of the American Historical Association for the Year 1931 (9638, 72 H.Doc. 317); and the very interesting sounding Brewing and Liquor Interests and German and Bolshevik Propaganda, (7597, 66 S.Doc. 62) as a start.

Another method that I used to learn about the types of content found in the Serial Set was to peruse bibliographies of books to see if the items I think should be cited from the Serial Set are, in fact, there. So a book on the transcontinental railroad should cite to the Pacific Railroad Surveys, a book on Pickett’s Charge should have a few references to the War of the Rebellion Papers, and so on.

When I got reference questions that I had absolutely no idea where to go, I would often ask the patron, “Which agency would be interested in publishing on that topic?” If they had done any reading or research before getting to me, they often had some interesting ideas and more background that I could draw from. Given that the government publishes on topics including dairy cows, mad cows, pesticides, epidemics, foreign relations and government scandals, if you think there is a chance that a government agency was interested in the topic at hand, it’s worth a search.

What Functionality Do You & Your Patrons Want?

When considering a digital purchase there are many factors to consider, including:

- Does the software allow you to easily search all the fields you expect in the document?
- Can you do both simple and complex searches with ease?
- Can you limit your search?
- Can you easily search for relevant text within the document?
- Can you search within the documents for specific illustrations, for statistical tables, or for maps?
- Can you and your faculty share the document with students or other researchers outside your institution to facilitate research and discussion?
- Does the interface help or impede your work? Can you easily move about the document, or do you need to endlessly click and scroll?
- Can you easily download the whole text?
• Do you have a “huh?” response when looking around the interface because something isn’t what you expected, or just the “aaaaah” of satisfaction because the interface behaves the way you expected it to?

HAVING IT ALL

As a reference librarian, I want it all because my patrons have different needs. I want the ability to easily and quickly search by report number, document number, bill number and even by SuDoc number. How often does someone come up with an incomplete or just plain weird number or citation that needs to be verified?

Often times, queries that made their way to me, as the federal documents librarian, had document or report numbers that didn’t exist for the Congress cited, or the information that got pulled up didn’t relate to what the patron thought was being cited. So I needed the ability to explore the various aspects of a citation – the report numbers, the Congress/session numbers, the titles etc., to get to the bottom of the hard questions.

I want to be able to do the tricky search and download the entire document and mail it off to a faculty member to put myself in the loop of his or her research [and make myself look good in the process!]. I want to assist faculty in putting documents on reserve, be it electronic reserves or the paper variety. With the ProQuest® Serial Set Digital Collection I can do those things.

SEARCHABLE PDF DOCUMENTS

I want the ability to facilitate research. The search interface and the document interface should make this easier, not harder. A good example was posted on GOVDOC-L (August 2005) about the report for S. 784 in the 40th Congress 3rd session. While the report was never published, I decided to look in the Journal of the Senate to see if there was information about what had happened to the bill and the report. A search revealed the Journal (Volume 1359) and I started downloading the 631 pages and searching. There were eleven “hits” on S. 784 in the Senate Journal.

Clicking through the PDF and reading about the progress of the bill, I was struck by how much easier it was to search the PDF for this information than the paper version. Usually with the paper version, I’d end up with all of my fingers holding pages (or I’d look around for paper scraps to mark the pages) as we referred back and forth from the index to the pages cited. According to the Journal, the bill (S. 784) was reported with an amendment; however there is no record in the Serial Set of the report, and no citation to the report other than the phrase “reported with an amendment.” So I learned the bill had been amended. And I was struck by the thought that there’s a reason PDF is a
default standard – it works so well for searching through the document; the PDF interface behaves in a way that facilitates the research. No more flipping through the document, going back and forth from the index to the referenced pages, but an easy click-click-click as I followed the progress of the bill. Unfortunately, back in the days of the paper world at the institution where I worked, the Journals were too big to be shelved near the Reference desk. They were either six floors away, or out in a storage facility page-able within 24 hours, but not available for a reference question. Having these basic resources available online is an incredible boost in the service I could provide across the desk, over the phone, or even in email.

**DURABLE URLS**

Further, the ability to share the document, either via a URL that doesn’t disappear (called “durable URLs” at ProQuest) or via downloading and emailing means that you can cite to the document from bibliographies, send the URL, the entire document or portions of it (depending on the sizes your respective emails will allow to be sent), to a student, colleague or faculty via email, or use it within the library for services like interlibrary loan or online bibliographies.

**24X7 ACCESS**

To me a search like the search for S. 784 (40-3) reveals the utility of a digital collection. We all know bill tracking is a steady part of the diet of a documents librarian, so having the digital collection available at the Reference desk, at your own desk, and even at home, is vital for the kind of work librarians are doing these days. In our twenty-four hour world, reference takes place with faculty and graduate students via e-mail as much as when you’re physically working at the reference desk. The ability to assist them by downloading and emailing, allowing faculty and assistants to post documents for e-reserves or on their own servers for their collaborative research, or simply pointing them to the correct document makes all of our lives easier.

**SEARCHING ACROSS COLLECTIONS**

Like many of my colleagues, I want the ability to search across databases for similar material as well. Sometimes this is enabled through services such as federated searching and sometimes it’s a matter of putting the right databases together. ProQuest is already working with vendors on federated searching. Users are also able to search across ProQuest® Congressional databases (the modules your institutions subscribes to) and find, for example, references to Ronald Reagan when he testified before the Committee on Un-American Activities, references to him in Committee Prints, documents from the Serial Set dating from when he was Governor of California (in the 1960s) about California parks, CRS reports from the 1990s that reference the Reagan
administration Energy Tax Policy, eulogies after he died, and more. In the old interface it takes at least five or six – or more – searches to mine this information, with the new interface, users can find all of this in *one* search.

**PORTABLE SEARCH RESULTS**

Along with the usability of searching the *Journals* shown in the example above, I want documents that are fully portable. Portability means users can download the entire document and exploit it however he or she cares to – take it on the bus, the plane, search, print, delete, use software to highlight parts of the document and comment for online courses, post documents for electronic reserves, or email copies. Users shouldn’t have to be tethered to our interface to use the document -- whatever is technically feasible to do with the document, I should be able to do it without impediment, and it’s possible; ProQuest doesn’t try to control the uses of the document.

**PAGE NUMBERS THAT MAKE SENSE**

One of the questions I’ve heard about the ProQuest® U.S. Serial Set Digital Collection has to do with the pagination. In our OPAC worlds, we’re used to seeing information in the bibliographic record about the number of text pages in the volume, and perhaps appendix or introduction pagination. The *Serial Set* is complex – many different kinds of documents and different kinds of pagination to go along with it. Bibliographically, legislative reports on bills are very often quite short, i.e. less than three pages and usually they do not have illustrations or maps. Consequently, these match up very well to the pagination of the PDF. The executive documents, however, are a different story. They’re the ones with the majority of pages, although they are a minority of the total number of publications. The documents can be quite long, sometimes many hundreds of pages, and sometimes these documents have multiple series of pagination. Multiple paginations within a single document or volume are particularly dicey -- some departmental annual reports show that each bureau report might have its own pagination, other times appendixes, attachments, exhibits, or enclosures might start the pagination afresh. So a given publication might have six or seven separate "page" 1 through 20s. Adding to this is that all illustration plates and maps are not included in numbered pagination, but they do add to the collation (total number of pages). Consequently, the editors determined that to provide the collation, rather than exact pagination, would help the user navigate through the PDF documents.

**SEARCHABLE TITLES OF ILLUSTRATIONS, MAPS & TABLES**

Being able to search the metadata for the illustration captions is another enhancement. While it doesn’t happen with great regularity for librarians, faculty
and students often want an image of a specific item – for example, the battlefield at Petersburg, or Eskimo costumes. They also may want a table with wages from the early part of the twentieth century. For example, the ProQuest® U.S. Serial Set Digital Collection doesn’t just provide the ability to search for a document with the relevant text; the metadata includes the image and table captions, thus allowing users to search for the needed illustration or table.

So the collation information in the metadata is especially helpful in using the citations to the maps, illustrations and tables and reflects the page number for the PDF. For example, searching the illustration captions for “Colombian berry” brings up one citation, *Yearbook of Agriculture Department, 1921* (8047, 67 H.Doc. 399). This has the caption metadata “Fig. 10.--The Colombian berry, a promising new fruit, introduced in 1921… (p. 32 in PDF).”

A search of table titles for “cost of living” brings up a number of hits, among them *Amend Central Intelligence Agency Retirement Act of 1964* (12792-5, 90 S.Rpt. 1504) with several tables charting cost-of-living adjustments on pages 2 and 3 of the PDF. This way, users know that going to the cited page in the PDF will have the illustration or table they searched for.

**ACCESS FOR SCHOLARS USING ASSISTIVE DEVICES**

Another function important to librarians today is that of ease of access to those individuals with disabilities as provided for in Section 508 of the Rehabilitation Act of 1973, as amended (29 U.S.C. 794d). While Section 508 the Rehabilitation Act deals with federal government procurement standards, in the spirit of accessibility ProQuest has met Section 508’s standards and has looked at both the search interface and the digital documents that were created for the Serial Set to make them as user friendly as possible to the visually challenged. With ProQuest digital products, the security on PDFs is such that underlying text can be read by screen readers. This means that users with vision problems can have the documents read to them by screen readers such as the popular JAWS® program.

(And I’ve heard this – the fellow who sits in the cubicle next to mine has his software set for a British accent, so I get to hear the documents read aloud in the Queen’s English).

The interface of ProQuest® Congressional also provides “alt” tags for all graphics for proper screen-reader interpretation. For clickable items “alt text” will contain descriptive text, for example, that the graphic is the corporate logo (users without the screen reader software can see these pop up when they mouse-over a graphic).
The text is also written and organized so that screen readers can easily read it so that the data will make sense when read (users without the screen reader software can check for this by removing style sheets and having the text appear in an organized fashion). “Skip” navigation such as that used in the ProQuest product can be employed to allow readers to skip repetitive portions (for example, the left navigation bar), and move directly to the searching portion of the page. Users without the screen reader software may inadvertently mouse-over the area where the text is hidden and see the command “Skip to the end of navigation.” Users can also check for this by looking at the source code and searching for the word “skip.” Because the screen readers read each word on the page, left to right, top to bottom, the skip navigation creates important shortcuts. While ProQuest was redesigning ProQuest Congressional, they also re-tested the interface and documents to make sure there are no surprises. There are ways to be compliant and still be screen reader unfriendly, and they want a compliant and friendly interface.

**HELP TEXT**

The help text is another area where CIS products have shone. In fact, customers have said they can put a novice reference person on the desk and, because of the quality of the help material, they can find their way through the complexities of the Serial Set. The Serial Set Index’s User Guide talks not just about the kind of documents users find in the Serial Set, but it also explains how they got included in the Serial Set and differences in material that has been included or excluded over time. This material has been moved to the online environment and updated and amended as necessary. Items you’ll find in the help text include instructions on creating durable URLs, tips for users searching the database and more.

**The Story of the ProQuest Digital Collection**

When ProQuest started the process of creating a digital Serial Set, there were already three building blocks in place – the CIS® Serial Set Index for the basic bibliographic records, the thesaurus from the Executive Branch Documents 1789-1909 for the backbone of the subject indexing (and which helped considerably in the redesigned ProQuest® Congressional, as well), and the CIS® Serial Set Microfiche collection.

**DOCUMENT SCANNING**

ProQuest used their Serial Set microfiche and scanned that material, converting it to digital documents. (These documents, with OCR, or optical character recognition, allow users to retrieve documents via full-text searches and search for terms within the documents). Given that they started with microfiche, not paper, they were able to digitize the material more quickly, and yet if there were a problem with either the image or
the OCR quality, they could adjust the settings and re-run the material with no additional wear and tear on a collection.

So that images aren’t a third generation scan from the microfiche, illustrations in the original paper copies are also reimaged. This insures availability of a high-end replica that could be used in books, journals, and other scholarly endeavors. This means that the Colombian berries are a high-end image that could be pasted into a paper on agriculture in the Andes Mountains, sepia images are sepia-colored, and the gorgeous color images of owls and woodpeckers, such as those found in the Report on the United States and Mexican Boundary Survey (834, 34 S.Exdoc. 108) almost fly off the page with their glowing colors.

Along with the re-imaging, ProQuest optimized the PDFs. What this means is that you’re downloading a smaller document. If you created a 17-page PDF on your desktop you might end up with a 3,700 KB document. Because the document is optimized before it gets put on the servers and delivered to you, the document will be less than 1,600 KB or less than half the size of the original. That means shorter downloading time, and a smaller document saved on your hard drive. A win-win situation.

INDEXING

While the CIS® Serial Set Index was lauded at the time it was created as state-of-the-art, the key-word-from-title index simply doesn’t work as a true subject index – there are various terms in the subject list such as “superintendent” and “revision” that don’t assist users in creating a focused search. Using the Executive Branch Documents 1789-1909 for the backbone of the thesaurus, and combining it with the other data from the CIS Serial Set Index, editors have developed a controlled vocabulary that reflects both the legislative and executive material found within the Serial Set. The editors have harmonized this indexing with the material found in the historic indexes and in the ProQuest® Congressional Research Digital Collection (which covers CRS reports and committee prints). It is this harmonized index that will make the subject search across the redesigned ProQuest® Congressional possible.

CIS products have long been known for the quality of the indexing – often the indexing in the products was what got me to the document I needed – there was no other way I was going to stumble across the answer. The overarching philosophy employed across the CIS products has been to index to the “mid-level.” Each product has its own rule about what “mid-
“level” means, but the goal is to have the main themes in the documents covered in the indexing. This philosophy ensures that the subject indexing doesn’t pull up overwhelming results, and that the results are on point.

For the **Serial Set** we’ve had the indexing team go through the material several times. The first pass through the data was to pay special attention to documents over one hundred pages (these documents make up 80 percent of the pages in the database!). The mid-level indexing for these items ensures that a search, say for “Amistad,” won’t bring up the references to US Postal Service stamps. In a similar scenario, someone’s great-grandfather may have fought in the Civil War, but that doesn’t mean he should be named in the subject index. A full-text search is the answer for the passing mentions of people and stamps. Our editorial systems also allow us to look at similar reports over time and thereby control the subject terms to ensure both comprehensiveness, and, where appropriate, consistency in the application of subject terms. Subsequent passes through the material were made to enhance the indexing of the smaller documents and to add SuDoc numbers, both to the main document and to those documents within documents which were unbundled. This brings me to the next enhancement.

While the review of larger documents was being done, editors also took the opportunity to “unbundle” documents such as the **State of the Union** addresses which, until approximately 1915, often contained other Executive Branch documents that were sent to Congress along with the address -- creating huge documents (with duplicative pagination). They separated the documents and gave each publication separate bibliographic data and indexing. An interesting side-benefit of the unbundling is that it allows users to see which documents were bundled together, and to compare that with the subject indexing of the address to see what the President was emphasizing, or deemphasizing, in his address that year.

There are notes in the help text about how to recreate the entire document, so if, for example, you run across the 1862 **State of the Union Address** (1156, 37 H.Exdoc. 1/1) you can search for document number “H.exdoc.1” in the 37th Congress. ProQuest® Congressional automatically truncates the document number so you’ll retrieve the sixteen additional documents that were originally bundled with the **State of the Union** address:


There are some instances where more than forty documents are now unbundled. The
fact that editors are taking them apart and applying subject indexing and the rest of the bibliographic metadata to each of these documents means the relevancy of your search results is increased. It also means that when reviewing the document, users aren’t forced to wade through potentially thousands of pages of text to find the needed document/section.

Along with the subject indexing and unbundling, editors have added SuDoc numbers based on past work with the Executive Branch Checklist, and also added section-level SuDoc numbers to documents that were not unbundled. The Annual Report of the Secretary of War, for example, often has other reports such as Report of Quartermaster General, the Report of Surgeon General, the Report of Adjutant General, and the Report of Commissary General of Subsistence, to name a few. As a result of this important work, there are more than 33,000 SuDoc numbers included in the indexing and, of course, to a documents librarian who speaks “SuDoc” this is a very valuable way to search the collection.

**Conclusion**

I hope I’ve been able to answer, at least in part, what that thing we call the *Serial Set* is, and how ProQuest has put together a product that meets your needs. If you’d like to take it for a test run, call us, or send an email. Give it a whirl – it’s a pretty cool thing, and it’s only going to help your users.

**End Notes**


3. Document Portability Note: 1194.22 Web-based Intranet and Internet Information and Applications and 1194.24 Video or Multimedia Products (see [www.section508.gov](http://www.section508.gov)).