

## Introducing the U.S. Congressional Serial Set Collection – Audio transcript

---

[ Please standby for realtime captions ]

good afternoon everyone, if you're just joining us, we will be going into our discussion about the U.S. Congressional serial set in 15 minutes. Thank you. hi, Heidi. Would you like to do an audio check?

---

Sure. Can you hear me?

---

Yes, you sound great. How about you, Suzanne?

---

I can wait.

---

That sounds good, too. Thank you.

---

Hi, everyone. If you're just joining us, we are five minutes away from our next session on the U.S. Congressional serial set collection. We will be with you soon. Thank you. Welcome back once again, everyone, and welcome to our session introducing the U.S. Congressional serial set collection. We have three speakers here tonight. James Sweany, Chris Ehrman, and Bailey DeSimone. Think of her joining us. We have two speakers, Heidi Ramos and Suzanne Ebanues. If you have any questions, please insert them in the chat box and we will get back to you at the end of the chat session with answers to your questions, and of course, a reminder that we are recording our entire conference, and we will release a recording to all registrants at the end. With that, I'd like to handed over to Susanna Ebanues to get us started.

---

Thank you, Kelly. And thank you for joining us for introducing the U.S. Congressional serial set, a joint PPO with the Library of Congress presentation and project. As Kelly mentioned, my name is Suzanne Ebanues, and I made supervisory management analyst. The other speakers this afternoon or Heidi Ramos, the lead program planner, James Sweany, chief digital resource division at the law library of Congress, Chris Ehrman, digital project coordinator an additional resource division at the law library of congress. And Bailey DeSimone, library technician of the metadata in the digital resource division. The project digitize the U.S. serial set is a project between the law Library of Congress and GPO. Were managing the digitization of volumes. We are conducting quality control digitization and then sharing the digitized volumes with GPO. GPO has been creating the metadata, designing and creating the collection, acquiring replacement volumes if needed. Initial public release, the collection was last announced on the 28th. It's now available from the library of congress website, and gov info. Why did the law library and GPO decide to take this product? And the simple answer is that it's amazing. It contains primary source material for a lot of American history. Is comprised of documents and reports of session starting with the 15th volume of Congress. Residential communications to Congress, such as the state of the union, trading materials, executive agency materials, reports on harbors, and other governmental publications, for example, the annual reports of the Boy Scouts. We are digitizing volumes from the 15th Congress through the 103rd Congress. Within the 15 through 103rd congresses, there are an estimated 15,850 volumes and around 12 million pages. So this project will continue for several more years. Such a massive amount of content, you need a large team of dedicated and hard-working people, and I wanted you to see the faces of the GPO team from library services contact management and

project strategy and technology made this project possible. These are the people who create the metadata, design and maintain the info, process incoming content, as well as locate replacement volumes. They would not be a set collection without them. Today, Heidi and I want to share some of the background processes behind the collection with you. In the preprocessing phase, we received digitize content delivered on hard drives, which are then transferred to the network. We violate hash values in a structure called Braggs, using the braggart tool. We then move into the metadata creation process, which can be a lengthy process. The metadata team either extracts marked records, or creates original records for all the individual roof words and records containing a serial set volume. This can range from one document to several hundred volume. Team members run a series of scripts that were run for this object to update the records and validate certain fields. Thanks to the efforts of the metadata team, all the records contain authorized Library of Congress subject headings in the fields. The team created a large effort last year to into authorized subject headings to allow better subject access to materials. They are now working on transforming other headings into authorized forms as well. The script is used to create an XML record. A nonmarket metadata volume for each file is created. This data is pulled from the delivery spec sheets and a master serial set spreadsheets. The master spreadsheet has information from the University of North Texas serial set inventory, the Congressional serial set inventory, which is maintained by Washington University. The College of Worchester, the 1909 checklist, the numerical's list, and schedule of volumes. I will now turn the presentation to production.

---

Thank you, Suzanne. Thank you, Suzanne. So as Suzanne said, at this point, the metadata is created, and the volumes for volumes and individual documents. And at this point, we kind of start the whole package creation and mission days of the process. And basically, what that means is these ready bags are transferred to the guv info processing area. And at this point, we can kind of pick up and rerun script that's going to validate, because again, that's a handover. We are going to validate that all the expected files are in there with in that bag. While the XML, the PDFs, all of these files, we need to ensure that everything is in there. And that's really important, because we want to make sure that we are not missing anything, and then it further validates that chain of custody from the content provider, in this case, the law library, all the way through. We can capture that information and premise. If we run through this process and the content is not valid, it gives us an opportunity to kind of -- It flags it, and lets us another some type of issue. We can go back and address that when it is valid, or if it is valid. It continues that process and the metadata are going to be converted into submission information packages, or what we refer to as sips in the government InfoWorld. Really what that means is, imagine that were getting this really big holder that contains volumes and individual PDFs. It has files, JPEG's, and all these various metadata files in heard of this one big bag. And what we are doing is converting those into those individually structured info packages. So each individual volume, each individual document is coming its own structure package. In those get uploaded to the guv info content management system. So once uploaded, then they are submitted. And at this point, that's really when the process except for file processing and parsing. For the serial set, this is going to mean that the volumes in the individual metadata -- Individual documents, metadata is parsed from XML files for each of those individual SIP's. The files are renamed off of access I.D.. PDFs are going to be individually signed, and then the content management system and preservation repository is updated with the packages. And then last but most important, this now means that we can actually publish the content out to our public site. So publishing out on a public site, that just basically means that now the public has access to all of the individual volumes, as well as individual documents on govinfo. This collection in particular has lots of search and browse functionality . There's the ability to search using basic and advanced search options. Users can also utilize collection specific search operators and filters. There are citations searching. We've incorporated the collection into the congressional documents and congressional reports collection. So you don't necessarily have to know that something is within the serial set. And then, also, you can

access the collection through govinfo's site maps. And I always love to promote this. If you need assistance, we have a really robust help page that you can go to to access to help you with your search strategies. So at this point, I've talked a little bit about, and showed you some slides. But what I really want to do is going to a live demo so you can actually see this for yourself. So I'm going to share. All right. So due to a little bit of our time constraint for the session, I am -- If you are interested in getting into the details of how to search and browse, I do recommend taking a look at that webinar from September 29th. But otherwise, I'm going to show you a couple of really cool searches I think some of those -- Those of you mentioned some were really interested in illustrations for the serial set, as well as maps. So I'm going to show you some quick ways using a basic search on how to act to stop. So just to start off, like I said, I'm going to jump right in, and I'm actually going to choose to use a collection search operator, which in this case, his collection cereals that. And then I'm just going to look for illustrations. So this is just a standard search results page for gov info. Take a look over here on the right. I thought this was a really cool way for some of you that are really interested in trying to dig down in to find some of those cool illustrations. Because we do have some images for you. You can kind of see some of the cool stuff directly from the search results. The other thing I want to point out while we are on the search results page, like I said, I specifically narrowed into serial set. But notice your collection, for those of you who want to refine your search or filter things, we do have these other collections within. And that's because these are actually part of the serial set. Other options for you over here is date published. Government author, Congress number, you can filter by serial number, committee, document class, president, organization person, topic, and place. Topic is really cool. When available, you have the ability to go ahead and sort A-Z. And I'm going to do that. And were going to take a little bit more of a closer look at these. So say I wanted -- I'm scrolling down, I'm saying, oh well. Animal fossils, that looks pretty cool. Let me check that out. I did pick this one specifically because it has this really cool color image. And while I am here, I wanted to talk a little bit about what your souls results look like. First line you are going to see years when to be a citation. So how's document 6-428. And then a title. The second line is letting you know that you are looking at the serial set, and the numbers 8607-1. Next to that will be the volume title, in this case, how's document volume 46, part one. And then a date. This line will also include committee, if that information is available. Below that is our standard teaser, and of course, I look for illustration, so that's why we have that in bold. And then, of course, access to PDF details and the ability to share that. I did download this particular one also, because I really quickly wanted to show you how really cool this is. If you happen to dive a little bit deeper in this one, there's really, really great images and illustrations found out this particular results. And I thought that was kind of cool. Lots of really cool stuff, especially as you go further in. Great. So that is sort of the search. We also have the ability to conduct an advanced search. And you have the ability to do a citation search. Before I jump into that, for those of you that love maps and I wanted to know about maps, I want to throw in this just so it is recorded. This is a really great search that searches across all the serial set. Specifically looking for just the volumes that contain maps or color maps. And this is one quick way for you to note that to find maps. Also, in addition, when you're looking at volumes, we want to jump into details because I do want to show you the mods for this. There is so much metadata that we were able to capture, as Suzanne pointed out, from various resources. So really take a good look. This particular one, as I pointed out, this is what it was searching across. It says it's true. There's one map, there's one colormap. It also has some foldout. You can scroll through some of the mods, especially with volume level, and you will see some of this really cool information that we were able to capture about this aerial content. So as I said, advanced searching is available. We jump over to that, so -- If you want to do an advanced search, just click over there to advanced searching. And of course, you got your drop-down of all the various things you can search within. My favorite is document number because I like to just throw a number in there, I don't always like to do Congress. He also have the ability for zero set volume, or a specific serial set volume number. Lots of great information. There is also the ability to conduct a citation search, so for example,

if I was really interested in congressional serial set, say a house document, I would choose the Congress. Say it is document 22. I search that, and of course, that will pull up -- As I mentioned earlier on, I also said that were possible, we associate the serial set with collections. I could do the same citation search using, say, congressional documents. I'm just changing the collection. I choose house document, six Congress, and document 22. And as you can see, that same search -- That same document was retrieved. So that's kind of what we mean by you can use, when possible, it's associated with the other collections as well. One of the final things that you could do with this collection is Rouse. You can either go to a to Z. But because we happen to be in content details, were going to cheat a little bit and go to content actions and browse. Quickly kind of show you what our brows looks like. It is broken down by Congress. Next to Congress, it does give you the Ranger serial numbers that are available within that con Congress. So if I'm going to take a closer look at this particular Congress, the next level down, it will be that serial set number. So if I was interested in serial number 8525, it's going to give me that volume title as well as a publication title, miscellaneous reports in the range of the various different words that are found in their. Going down and clicking in further, it's just -- It's exactly what it is. It's a browse tree. It's going to break down for you everything that is been in that particular serial set. So I can either look at the entire publication, or I can jump down and kind of browse through all of those really cool documents that you can find within that particular volume. For those volumes that have a significant number of individual documents, and we all know that Congress likes to bind quite a few of them together. You also have this ability for you to go through pages. So you can reset your page limit up to 500, or you can scroll through by page route. And as it was announced, a couple times this morning, as well as it went out as an article, we have released our first enhancement for the serial set. And because I'm here on this Rouse page, I'm going to jump and kind of show you how you can access that from here. There's also the ability to access it from the brows, and that's a congressional serial set browse by topic. This brows basically is just taking the serial set documents and breaking them down into topical categories. These are all based off of the Library of Congress is subject headings. The brows is an A-Z index. You just have to go take a look, find a term, or a topic that you might be interested in, by clicking on the various letters. I have to be honest, I have seen pictures on Twitter of cookies and baked goods, so my mind is on bakeries. So I wanted to find out, hey, what are all the cool things available in the serial set under this topic? I would just have to click on that, and is just a breakdown of all the individual documents that have that particular topic in the metadata. So the topics on display for the serial set documents are what is currently in govinfo. So each individual document is assigned one or more topics . And just to let you know, as more content is brought into the collection, you may see additional topics as well as the topics available that will expand. So that's just one of the new features. We really hope it makes it easier for features to explore the really rich information you can find in the serial set. And with that, I am going to stop sharing and bring us back over and handed off to Suzanne, who is going to talk a little bit more about what is next to come.

---

Thanks, Heidi. We hope you will all have a chance to check out the collection. As Lori mentioned earlier, we are continuing to work on the enhancements. Additional volumes for the serials that will be added to the collection as they are digitized and have complete metadata. We have also started planning a series of short collection tutorial videos that Mark records for individual documents and or sitter readily available on govinfo , they will also be available on the catalog of U.S. Government publications. As we look to make more enhancements and modification as to the collection, we want to hear from you. So please share your feedback on the collection by using the link on the light. And now it's time to hear from our partners of the Library of Congress. I will hand the ball off to James.

---

Thank you, Suzanne. And good afternoon. Again, I'm James Sweany of the resources division of the whole Library of Congress . By colleagues, Chris Ehrman, Bailey DeSimone and I will discuss the work of

the law library to discuss the congressional serial set in collaboration with the U.S. government publishing office, which has begun its third year. Again, to recap, this project is a multiyear effort to digitize and make accessible volumes of U.S. congressional's serial set back to the first volume, published in 1817 by the 15th Congress. Through 1995, or the 103rd Congress. Again, the law library is inventory and digitizing the serial set, and the GPO's cataloging these volumes and authenticating the digital files. Library of Congress will display, and is displaying, the serial set volumes for Republic access at the Library of Congress website, and the serial set documents will also be displayed on Congress.gov. The digitization is expected to take about five years to complete. For fiscal year 21, the annual performance goals for the serial position process in the law library were, one, to prepare digitize 3000 volumes for digitization, and publish a subset as a digital collection displayed on the website. Since October 1st, 2020, the law library's serial set team reviewed serial set volumes for completeness and condition, tracked volumes through the review process, and prepared 3800 Avenue volumes, shipping 3349 volumes to the vendor. A subset of the serial set, 287 volumes of the 69th Congress was launched as a digital collection on the law library's website. Now since the serial set project launched at the beginning of fiscal year 19, over 6000 serial set volumes have been evaluated and selected for digitization. So we are approaching 40% of the total volumes. 5700 volumes were sent to the vendor for digitization. We received back about 4300. And of these so far, 3800 have passed the quality review, and we are sending to GPO these volumes that are passed quality review. And so far, we have sent 1500 digitize volumes to our partners at GPO. In fiscal year 22, the law library is preparing an additional 3000 volumes of the serials that for digit station. And so we will proceed forward. Now, to discuss her evaluation a digitization processes, I turned to my colleague Chris.

---

Chris, your mic is muted.

---

Sorry about that. So within -- I'm going to talk about the -- The Library of Congress. So within the Library of Congress's collection, we have multiple copies of each serial set volume, and we inspect these volumes for completeness and any physical characteristics for documents that are -- Like Suzanne mentioned, and also the Worcester college resources. We record the physical characteristics of the -- And if a volume is complete, great. We selected for digitization. If not, we request some of the -- Or request volumes for GPO. In some cases, we started creating combination volumes for multiples serial set numbers. There's an access database that we called the STP. It sounds for serial digitization project. The project tracking database allowed us to combine all our sources, such as the 1909 index. And also record the physical characteristics of the volumes that were evaluated and the volume that was selected to be sent to our digitization vendor. So we ship the volumes off to our vendor, and they digitize the volumes to three start value performance. We receive a master JPEG, and an OCR XML file. A thumbnail that is a gift for each page, and then a document level PDF, and a volume level PDF. Each serial set volume has a unique tag I.D., using the serial set number and in part and volume. And then we also use this unique bag I.D. for the volume P BF filename. We have document directories, which include the sequence of where the document appears within the volume. The document number, and possibly part as it is printed on the page. So this gives each document a unique I.D. that we can -- And so these are delivered to us by our digitization vendor and a bag structure via STP. Our institutional repository is the content transfer services. We use this to receive bags and start workflows. As Heidi said, we -- The bags have a wrapper that describes -- Has checksums for all of the files, and we verify that nothing has been lost in the transfer from the vendor to our systems. And after it is accepted, it's copy to long-term storage and also public access. So, to evaluate the digitization, we use in-house created software called sampler. And it allows us from quality review of the images, for the images, we look at the thumbnails. We check that all the documents are present and are expected, and also for completeness. And then, the STP project tracking database also allows us to display the information about the digital object. So

we have the information from the 1909 index is, UMT, and then we also have the additional metadata that our digitization vendor provides us, such as number of holdouts, maps, number of documents. So we are also able to record the status. Like, if we found the volume to be complete, we accepted. If there was an issue, such as a volume level PDF that was originally -- As we receive rework from our vendor. The STP also allows us to create visualizations using tableau. We just link it to the backend of the database. The snapshot as of year-end, we are able to keep track and easily see how many volumes have been polled, how many copies have been selected, how many digitized bags we received and accepted back, and it's just useful for reporting out. So, next up, after is accepted, I utilize CTS. It has an API to export the bags for delivery to GPO for authentication and cataloging. And then the next steps for working on establishing method of using the API to cover the authenticated PDFs into our library systems, and that additional functionality, such as more search functionality. I will pass the presentation back to James, who will talk about the law.gov website and the presentation. Thank you, Chris. Now, to place some perspective, the presentation of the serial set on the library's website, I need to give you a little background on the refresh that has taken place on the law.gov site. And so the law library, working with the libraries office of the chief information office, we launched a comprehensive overhaul and modernization of the law library web site, which was launched on June 30th of this year. A complete redesign effort, including updates to thousands of pages that have previously been maintained by hand. A migration of content into the project application to allow law staff to take advantage of the advanced capabilities and complete the friendly design. It uses a new research center templates, which presents easier to explore layout and common components, such as featured blog post and seamless connections to other library pages. Multiple groups of digitized primary source materials have been converted to digital collections in the collections application, allowing users to search, view, browse and use digitized items using the same sophisticated techniques available to other GOP collections. Additionally, the law library's extensive reports collection has been updated to include new browsing capabilities, creating enhanced access to one of the law library's key product. A primary goal of the redesigned web site was to improve the discoverability of the law libraries collections. And more about the redesign is found on the law library blog. Collections that were migrated into the libraries video collections template included the National Transportation Safety Board, the Native American cost editions and legal materials, U.S. treaties, statutes at large, digital library, global gateways collection, and some of the rare collections. John Adams, and files. So it's a little bit of context, and we are now adding the U.S. Congressional serial set. With the refresh of the law.gov website, volumes in the serial set have been this late in the libraries digital collection templates, or is referred to as project one, and available to the law library's website law.gov. The law library digital collections in this template are synchronized. With the libraries general digital collections, and are also accessible from the Library of Congress website , LOC.GOP. With synchronization, the serial set display is available through the libraries digital collections, and law.gov. The interface is the same. In the about the collections section, related resources consisting of indexes on the list that we have used in the inventory process, as well as blogs and story maps promoting the serial set are found, which Bailey will discuss in a few moments. Presently, we are providing access to a volume level. However, we are developing access as well, a forthcoming enhancement. And now, I turned to my colleague Bailey for documentation and our promotional activities.

---

Thank you so much, Jay. My name is Bailey, and as we mentioned earlier, I am a library technician for me the data with the Library of Congress. So I do a lot of work with the documentation of our workflows, and tracking the progress that we have, and making the changes adaptable throughout the digitization process. So we currently track our progress in three workflow ways. These are documents that are continuously updated according to workflow changes and changes in the collection. We reflect these changes in an instructional guide for documentation to review what works, what didn't work, what we might alter the future, and so on. The digitization guide gives background information on the

serial set of the collection, or the foundation for the metadata that we create and track during the evaluation process. Physical volume evaluation and workflows are uplink and shared with library and contract staff involved with the project. The same for our QR workflow guide, it outlines the collection guide for digital volumes received from our vendor, which is Chris mentioned, is under a quality review process. Instructions are included along with specific workflow to evaluate each digitized volume. Formatting examples are included in the sample QR exemption guide. Instructions for documenting and formatting metadata are also included in each guide. Now regularly updated in order to reflect development in the serial set of the decade. Tracking metadata is especially important, because it allows us to create a short word bank of terms and raises that will help you filter your query searches. We are open with our internal staff, contracts, and vendors about the forming of metadata, as communicating volumes civic issues relies upon our ability to track metadata in a clean and efficient way. This especially applies for printed digitization errors that we record to determine that a volume is complete or accessible for public use. Metadata, both qualitative and quantitative. Volumes also contains color content like we saw earlier that needs to be treated with actual attention. To accommodate, we've identify these materials and developed formatting guidelines for describing counting them. Two of us ensure our volumes completion and help us identify purchase interesting concept information for public use. Apologies about that. Because a significant portion of digitization requires physical access to volumes, our team is adopted priorities and asked her to project during a time of extended Intel work. As we gradually return on site, combined volumes on in order to complete a single -- To complete a single volume. And any other volumes were donated. We prioritize the digital QR to provide feedback for our vendors, and also ship it our focus to the effort. Our contact staff are back on site as well, and we've been making efforts to communicate and collaborate to make sure that we are on the same page. One of our outreach efforts is through the law libraries log since January of last year. I made an effort to put at least one interesting or historically relevant documents from the serial set to the blog, along with some general contacts for folks new to legislative research where the collection itself, and you can be on the lookout for one later this month. We are hoping to feature the 69th Congress that we have just uploaded on law.gov. And one of the other projects that I've been working on that has incorporated the serial set content is story maps. So I also work with our summer intern, one of my colleagues, Jennifer Gonzalez. And what the story maps project does is it gives us an interaction with our online collection and it aids legal research and it frequently comes up. Right now, we have one published on the focused serial set, which is called city centers and defenses. You can look forward to many more from the serial set as we continued digitizing these along with GPO and our vendors. We are very excited to make some of these colorful images and maps and really interesting tidbits of information a bit more known in this way. All right, and that's all that I have. I can hand it back over to Ashley for questions.

---

Thanks, Bailey, and thanks everyone from the law library for that great presentation. I'm having router issues, so if you lose me, that's why. I'm not ignoring you. And I can't see any more of the chat, but I will be happy to have -- Thank you for joining us today, and we are happy to have any questions.

---

Okay. We only have about four minutes left in our scheduled session, so we will try to get through some of these. Earlier on in the program, Jenny Groom said just to clarify, you're talking about a record for an individual document and the serial set volume, too? The Mac that is correct. If you are talking from a CPA keep perspective, if it's a cereal, there's only going to be one record.

---

Okay. Jenny also asked, will search by serial set volume citation be added?

---

So there is the ability to search by serial set volume citation. It's not necessarily within that citation search functionality Amma but we have actually search operators that will do that, and is also the capability within the advanced search, and there's a great, I believe, example of it in the help as well.

---

Okay. Do we have any plans for making this collection accessible to the sight impaired?

---

The PDFs all are -- The PDFs are OCR.

---

Okay. So we know about the copyright on this, but many in the general public will not. I haven't seen many statements about reproducing images, et cetera, on government info. I teach people to look for this information from those digital items. Any plans for help to the general public for this? We want students to develop good habits.

---

Heidi, feel free to jump in. There is blanket language on the info that could have copyright.

---

Yes. I believe it's actually under the about section, there's an entire section about copyright information.

---

How does this compare to one available from commercial vendors in terms of context, coverage, and search and browse functionality ?

---

J, feel free to jump in. In terms of content, we are only just getting started. So I think you heard the law library team say is going to take six years just to digitize the content. It will take us just a little bit longer than that to get everything up there. I'm going to be honest, we didn't really look at the commercial -- We are not in competition with the commercial sites, so we built our site based on what we thought our U.S. users as well as our public users with need.

---

And for the law library, we do consult commercial sites just to make sure as we take the inventory, if we have a complete volume or to ensure that if we have missing volumes, we are able to consult about the copies. However, we are building our access base on the library's digital collection templates. And with this collection specifically, on how our research would want to use a collection. We are using this to get suggestions, but we are building it around our readers.

---

We are out of time, but I can fit one more question. What is the error rate for the OCR attempt?

---

So that varies widely between the volumes. The earlier volumes, the paper is quite different. It's really dark, and it's -- The text in there can be water damaged. So it's definitely less at the earlier side, but I saw that question looked at, and it looks like some of it was about 50%. But I would say it generally does a good job. A later volume I checked had a confidence level in the 90s. So it definitely depends on the volume, when it was published, and what it contains. Like, if it's a table, it doesn't do a very good job recognizing the contents of that sometimes, but generally I would say the OCR is pretty good.

---

Okay. I do apologize, we did not get to all the questions. It was a long list. But I will make sure that all of the presenters get a copy of all of the chat and the questions that came in, and we will share it with the law library team, and GPO. Thank you to all of our wonderful presenters. Thank you for a wonderfully informative session. And we will be back with everyone tomorrow at noon Eastern time. Thanks for joining us. [ Event concluded ]