Association of Centers for the Study of Congress
Annual Meeting 2017
Session 4: Library of Congress Manuscripts Division, May 11, 2017

Karen Paul (KP):

The Library of Congress provides leadership and sets standards in numerous areas of librarianship, preservation, management, and access. This morning, we are going to see firsthand how the Manuscript Division builds on this tradition of excellence. Established in 1897 as one of the several departments when the Library moved from the United States Capitol, its staff of four assumed custody of a collection of 25,000 manuscripts, which had accumulated during the course of the nineteenth century. This early concentration on political, military, and diplomatic history broadened after World War II to include cultural, scientific, and non-governmental collections.

Current holdings number nearly 60 million items, contained in 11,000 collections. They include some of the greatest manuscript treasures of American history. Foremost among the holdings are the twenty-three groups of presidential papers, ranging in time from George Washington to Calvin Coolidge; papers of approximately half of the individuals who have served as secretary of state; other Cabinet officers; members of the Federal Judiciary; chief justices; and military collections ranging in time from John Paul Jones to George Patton and Alexander Haig. Of the five senators judged by the Senate Select Committee in 1959 as the most outstanding in the history of that body, the division holds the papers of four: Henry Clay, Robert La Follette, Robert Taft, and Daniel Webster. Remarkably, the papers of members of Congress comprise approximately 900 collections in the division, making the Manuscript Division truly a center for the study of Congress.

So I'm going to introduce our speakers in the order in which they will present, and have asked each to speak for about ten minutes, and we'll have questions and discussion following their presentation. Connie Cartledge is senior archives specialist in the preparation section. She began her career at the Library in 1987. She holds an MA in American history from Mississippi State and an MSLS from the University of North Carolina. She has worked with many large, complex collections, including those of Jack Kemp and Daniel Moynihan. She currently is leading a team processing the papers of Anthony Lewis, a reporter and columnist for the *New York Times*, from 1955 to 2001.

Kathleen O'Neill is a senior archives specialist in the Manuscript Division. She earned an MLS with a concentration in archives and records management from the Queens College Graduate School of Library and Information Studies, and has been with the division since 2009.

And Dr. Sahr Conway-Lanz is a historian in the library's Manuscript Division, responsible for the division's U.S. foreign policy, military, and international history collections from the twentieth and twenty-first centuries. Before coming to the library, he was a senior archivist for American diplomacy at Yale University, and previously worked at the National Archives and Records Administration with the Nixon presidential materials.

Ryan Reft is a modern U.S. historian in the Manuscript Division and oversees the twentieth and twenty-first century collections regarding domestic politics, law, journalism, and culture, and he has been with the division for two years.

So, we'll begin with Connie.

Connie Cartledge (CC):

Good morning. Thank you, Karen; I appreciate that introduction. I'm going to talk about how we process congressional papers at the Library of Congress. In my brief time, I'm going to focus on two areas. I'm going to talk about the importance of initial planning at the beginning of a project, and I'm going to discuss how we manage the physical processing of large congressional collections. I will be using examples from two collections, the papers of Daniel P. Moynihan, senator from New York from 1977 to 2001; and Jack Kemp, a member of the House of Representatives who represented Buffalo and Western New York from 1971 to 1989. As Karen was talking about all the treasures we that we have here at the Manuscript Division, I was thinking these two collections also have treasures, but they also have some kind of unusual things, like football helmets, shovels, bulletproof vests. So these have treasures and unusual things.

One of the most important steps in processing a congressional collection is planning, and we begin by gathering biographical information about a person. We try to determine what issues are important to him or her as a senator or a congressman. We look at their staff list. We look at the committees that they served on. All of this helps us in determining what are the issues that are most important to that member or senator.

We also begin by looking not only at their congressional career, but we also look at earlier points of their life as well. I find it helpful to create a biographical note even at this stage. I mean, you do that for – you finish finding it, but I find it helpful. Here's an example of Jack Kemp's biographical note, and you'll see – I'm sorry, it's a little small. I hope you can see it. I knew he was a football player, but I didn't realize he actually played for as long as he did.

Next, we gather administrative material about the collection. We review the instrument of gift. We review catalog and accession information to get the number of boxes and their location, and we check to see if any box list came with the collection. After those steps, we begin to unpack the boxes and to survey the collection. Here's a photo of the Moynihan papers, unprocessed. Actually, this is a deceptive photo. We had actually removed some boxes and had begun processing, but we didn't have room for the collection in our stacks, so this is actually the cloak room of the Madison Building that we had to house part of the collection in. It actually kind of was throughout our division. And then the Jack Kemp papers were smaller, and you can see we have some oversized boxes as well as oversized items and record cartons.

It's often thought that a collection that comes from an office is in perfect order. That's often not the case. If you didn't know, I'm just saying. Since about the mid-1990s, collections seem to come less organized to us than in the past, and in discussions with my colleagues, we think that's for a couple of reasons. One, less money is spent on staff to do administrative filing, and two, the use of the personal computer has made it easy for people to produce files, digital and paper. In our experience, most collections are actually a combination. Some materials are in good order, others have no order at all, and then sometimes, a series may be a combination of both.

When you survey a collection to determine its contents, there are several things to keep in mind. Is there an original order? Are there gaps that indicate items are missing? Think about potential series, and by that I mean types of material specific to an individual's life or career. For instance, in this case, you're looking at senatorial papers, and in the case of Jack Kemp, House of Representative papers. For the survey of the Moynihan papers, I recorded the accession number of the box, came up with broad series categories and types of material. This is the preliminary form that we used. When I was in the middle of this, I realized we should have left more room for notes, but what I did was, frequently, sometimes a box could be half a page or maybe an entire page. You would just mark the series that were in a box and you would put the notes on it. We put this information into a database and created an overall list of the boxes' series by accession number. A spreadsheet was used to record similar information for a survey of the Jack Kemp papers. Here's a look at that survey sheet.

Compiling a series list for these two collections allowed me to see the type of material and the size of each series, and there are two major factors that go into how we organize a collection here at the Library of Congress. We take into account research value and how we most efficiently can utilize our staff time in processing it. To help us determine this, an archivist will write a processing proposal that's submitted to our supervisors and to the historical specialist who serves as curator for the collection. In this case, it's Ryan. We do this for almost all our collections. The proposal outlines an organizational scheme and recommendations for processing, and highlights issues that need to be resolved. As you can see, we have a little list that we go down, and it expands as you need to. For the proposed series, sometimes that can take a page or two on its own.

All series are not created equal. Each series will be processed according to its potential research value. For example, the personal correspondence of a congressman will be processed more carefully than, say, a clippings file in the press series. In this proposal, you will indicate what type of series processing you're going to do. Will it be box? Are you going to do series level? Are you going to do folder level? Or it sometimes may be item level, if you have to. And you address preservation concerns, including if it's possible to keep the folders for a collection. The proposal is going to be your roadmap on how you organize the collection.

The second most important element that I think is managing the overall physical processing of a collection. Usually, at the library, we process large collections in teams. A processing team consists of a lead archivist for a project, who is assisted by archival technicians or interns. If a collection is extremely large or has a very short processing deadline, a team may include other archivists. It's the lead archivist who does the survey and writes the proposal. The lead archivist is responsible for managing the day-to-day work of the project. They must develop a work plan to determine the sequence of processing, and several factors influence this: The size of the collection, the amount of space you have, how much staff assistance you will have, the experience level of the staff, the timing of the staff assistance, and the difficulty involved in processing a series. The Moynihan papers consisted of approximately 1,688 record cartons, about 1,553,000 items. The Kemp papers consisted of 584 record cartons, 25 oversized boxes, and 64 flat, oversized items, around 149,000 items.

I was the lead archivist for both the Kemp and Moynihan papers, and with both projects we had summer interns coming in about three or four weeks, so I had to decide where we were going to start processing. The key factors were table space and difficulty of the series. At the beginning of the Moynihan project, I think I had nine or ten tables that would accommodate a total of 135 or 140 boxes, which was not much when you had over 1,000. Here's a slide of the work tables. This is where I usually work. As you can see, it's unprocessed material, typical, and you have flags indicating what the boxes are.

From my survey, I knew that 70 percent of the Moynihan papers were comprised of legislative files that were in poor order. I knew that it would require a great deal of time and effort to arrange these files. I didn't have the space and time to tackle that, so I had the interns begin with the administrative assistant and chief of staff files and form replies in the constituent services series. These files were in good order and required very little arrangement, and only required a small amount of re-foldering. I think it was a good decision. The files revealed what issues were important to the office and provided an overview of how the office categorized legislative issues. Here's an outline for the Moynihan papers, and you notice it says part two. Part one was his pre-Senate career, which had been processed earlier, which was around 450 document boxes.

After processing the constituent services series, we then processed the personal office files, the press files, speeches and writings, and finally the legislative file. The Kemp and Moynihan collections were processed to the folder level. We left the majority of the folders as they were and did not put items in chronological order. Personal correspondence files of Moynihan and Kemp, and some legislative files that were of particular significance to them, were put in chronological order. With both the Kemp and Moynihan papers, archivists or technicians were assigned a series or a large legislative topic. They would analyze the files, decide on a plan of arrangement, and discuss the arrangement with me or provide me a written outline. Then they arranged their series or topic and provided me with a computer-generated list. And then I would review that list and make sure that it was consistent with our arrangement scheme and consistent in the terms that we use for folder headings. And then I would take that list and put it into an overall document that was our draft container list throughout the project.

Although we try to utilize the folder titles provided, sometimes they just weren't possible. I mean, I know you've seen that, those great folder titles, "Pending," "To Be Filed Later," "?," "Unidentified Rock," you know, these type of interesting things. So in that case, we would try to use terms and categories from their legislative issue list and their form replies, and we also used Library of Congress subject headings when we thought were appropriate.

So, to sum up in my short amount of time here, I'd like to say I think it's so important to make sure you take the time to plan before you start on one of these collections, and as you're going along, to keep track and manage carefully these projects. And at the end of the project, you have a process collection and a finding aid available for research use. This is my poor effort at artistic – this is a completed box from Jack Kemp, and one of the binders of the paper finding aids for Senator Moynihan. And also, I would be remiss if I didn't say that I'm fortunate that I have a talented number of colleagues here at the library, and much of the credit for the success of these projects go to them.

After processing the Kemp papers, we wound up with 395 document cases, around 118,500 items. The Moynihan papers comprised 3,741 document cases, 1,306,400 items. And then this is like, one of the happiest days you'll ever see. (Laughter.) It's like, these are only temporary labels, but we didn't care. We did get the finished labels on not too long after that, but that was a great day. If you're interested in looking at the finding aids for the Kemp and Moynihan papers, or other congressional collections here in the Manuscript Division, they're available on our website. Thank you.

KP: Thank you. Kathleen.

Kathleen O'Neill (KO):

Good morning. The title of my presentation in the program is supposed to be "Processing and Preserving Born-Digital Collection Material," but that topic is entirely too broad to address in ten minutes, so I thought instead I would talk to you today about processing and using born-digital collection material, specifically how the inherent qualities and challenges of born-digital material impact processing workflows, and in turn, how the actions taken during processing impact researchers in the reading room.

But first, I want to give you a brief overview of the Manuscript Division's born-digital holdings. We have approximately eighty collections with a born-digital component. I'm sure we have many more media in unprocessed collections, but this is what we've identified at this point. Approximately forty collections are available for research, and if they're available for research, that means they've been processed, they've been described with the finding aids that are available online and findable via searches. Right now, that access means onsite in the reading room. We do not serve unprocessed born-digital material.

The born-digital material in the collections span about 1986 is our earliest files, to the present. The size of born-digital material in the collections, we have a collection that just has one file, and we have another collection that has 1.8 million files. So, you can see the range of material that we have to deal with and finding a workflow that would work for something that's just one file and 1.8 million files was challenging.

As far as our congressional collections, Connie mentioned there's a born-digital component in the Jack Kemp papers and the Daniel Moynihan papers, and both of those are available to researchers in the reading room. As far as public policy-focused collections, there's approximately twenty. Those include the Frances Oldham Kelsey papers, the National Federation of Democratic Women records, People for the American Way records, and the Russell E. Train papers. In a large part, the digital content in these collections is secondary to the paper. The born-digital materials in these approximate twenty collections were created in the late 1980s to the late 1990s at a time when the computer was more used like a glorified typewriter, so often, the digital content is not unique in that it has been printed out and exists in the paper portion of the collection. These collections tend to have simple file structures and consist primarily of text documents and image files.

We are currently processing the American Lands Alliance records, which American Lands Alliance is a now-defunct environmental lobbying nonprofit, and the William E. Odom papers. Odom was a U.S. Army general, a former director of the NSA, and an outspoken critic of the Iraq War. These two collections I'd like to talk about, because they really represent a shift in the types of material we're receiving. Odom's spans from the late 1980s until the early 2000s, but American Lands Alliance, in particular, there's a paper portion of the collection that goes up to 1999, and then really, the record in paper form drops off, and the organization basically became entirely digital. So this is one of the first times where we have a collection where the born-digital component is as important as the paper, has as much research importance, if not more, than the paper portion. And in terms of the Odom papers, a lot of the runs of material, like diaries, they are completed in the born-digital portion of the papers, so this is really a real shift. But even more importantly, both of these collections have born-digital material that is unique material; it does not exist in paper form. And, in addition, these are large collections – they're in the tens of thousands, hundreds of thousands in one case – and they're complex, hierarchically and in terms of the file structures.

So, collections from the early days of computing have the challenges of obsolete file and media formats, and all digital content is subject to file and media obsolescence, and they can be easily, intentionally or unintentionally, altered. So documentation of the file metadata, the chain of custody, and processes on the files is much more important for digital content than it is for paper, although quite important for paper, too. But the more modern collections – I'd classify that as post-2000 – they have additional challenges in terms of scale, which we addressed, 1.8 million files. These collections tend to be large. Where previously, we were dealing with hundreds or maybe thousands of files, now we're dealing with hundreds of thousands of files.

PII, personally identifiable information, is a really big concern. It's scattered throughout these files. And, as I mentioned briefly, you have complex interrelated files. You have email databases. And then, again, particularly for congressional collections, donors seem to have a greater concern about PII and restrictions with digital material – that's possibly because of all the news about hacking and emails – and so you could be facing, with these collections, longer restrictions or tighter restrictions on portions of the born-digital part.

So we have these large hybrid collections and their related challenges. How did we decide to process these materials? Well, given the scope of the materials, we really felt we had to find the most efficient and sufficient way of getting the materials to the researcher, and properly preserving them in the process. When you're appraising them, you deal with a lot of these issues. We have digital forensic tools that can analyze and index files and create reports for your appraisal. You can do reports on PII, so you can pick up Social Security numbers, credit card numbers, finding where the email is in these files, if there's any classified information, which could be very important for congressional collections, restricting materials, keyword lists, and finding out how many duplicate files you have. But use your best judgment, because it can give you a tsunami of information. Beware.

For the Odom collection, we had over 55,000 hits for duplicates, 24,000 hits for PII. Now, of course, a lot of these hits are false positives or duplicate hits because there's duplicate files, but you sort of have to triage and figure out where your effort is best placed. So when we looked at that, then we said it's not really worth it for us to go through and de-dupe, and there is some value in the fact of seeing these dupes, to see how people created these records. Now, we're lucky in that a lot of these files in themselves are small. If we were dealing with duplicates of huge TIFs, we'd have to consider space. But we, at this point, don't have that problem.

So, given the scale and complexity, what choices did we make? Well, original order rules. Of course, archivists really appreciate original order anyway, but given that we have concerns about authenticity of the files and the fixity of the files, we wanted to find a way to move them from ingest, through long-term storage, to the researcher, basically messing with them as little as possible, because we want to be able to say these files exist in the form that they were given to us. We can’t vouch for how they were created, but this is how they came to us, and they have not changed since we received them, or if they did, we’ll be able to say we changed them for this reason. It could be a technical reason, it could be a redaction for PII, those kind of things. We don’t rename, in general, files or folders. We don’t de-dupe, and we don’t migrate files to preservation formats.

Now, there’s asterisks next to all of them, because of course, nothing is ever really straightforward. So we don’t rearrange because it supports provenance and authenticity. It provides a researcher with a true picture of the original order. But the exception would be transfers that don’t represent original order. A lot of times, particularly with congressional collections or nonprofit organizations, we’re getting files off a network or an archival space that they have, that they’ve already moved it from its original location. Sometimes, when that happens, things are clearly dragged into the wrong folder, so in those cases, we would change that.

In terms of renaming files and folders, that’s really more of a technical question, if your repository can handle special characters, foreign language characters. Our repository development center has written software to be able to handle the majority of special characters, and so they support our decision not to rename files. And even more importantly, a lot of these files are interoperable, interrelated, so if you have a database and you change a file name because it has an exclamation point in it, then the database is not going to work because it's looking for a file that no longer has that name.

As far as de-duping, because it would really take too much manpower and time for us to go in individually and assess each duplicate file, we de-dupe on what we call a bag level. So when we get our media in, we use the bagger tool to copy the files off onto our long-term storage, and so each separate piece of media is in one bag together by itself. And there are cases where we've gone through and found that one entire set of files on a media exists someplace else, and so that might be a case where archivists might make the argument that we don't need to keep that one bag, because they exist in their entirety in another bag.

Migrating files, we certainly do not migrate the simple text files. We will migrate audio/visual files, because if we can't access them now, we can't appraise them. And of course, email has to be migrated, and so the more complex file types we may migrate. But if you think about migrating to preservation formats, trying to do that for 1.8 million files, I had a technical person, when I asked about the possibilities of that, looked at me and said, "Who's going to QA that?" And I said, "Good point. We don't have the manpower to do that."

And because we don't rearrange the files, the finding aid sort of has a different function. It doesn't mirror the physical organization of the files. That's actually where the intellectual arrangement of the files happens. It happens in the finding aid. The other choice for larger files, or if you don't have the ability or time to do that, you can use file directory listings instead of a finding aid, or as an addition to a finding aid. It's very similar to an item inventory. It's key word searchable. It provides file names, creation dates, and in the case of some of our collections, it actually acts as an index for the paper. So, where you have a run of correspondence that might be chronological, it's very hard to find correspondence from a particular person you're looking for. But if you had a file index, you could possibly search, depending on how the files are named, for the material you're looking for or a particular date you're looking for.

So, how does this impact the researcher? Well, as I said before, what's on the page in the finding aid is not what you'll encounter in the files. Materials are not going to be collocated together, so you may have to request, in the case of the Library of Congress, more bags, because the material is going to be spread out. But more importantly, the computing habits and even work computing habits are really idiosyncratic and messy, kind of what Connie had spoke about with the Moynihan papers. And so, researchers must understand that it's not going to be as easy to find materials they're looking for without having some digital skills. Does that mean you need to be a coder? No, but you need to be comfortable using tools, coming up with lists of subject names, topics that you're interested in. This is actually something that we ask the archivists to do.

In general, since we're still working with hybrid collections, we always process the paper first, because the paper informs the digital content. That gives the archivists an opportunity to come up with, if they've discovered Social Security numbers. You can start making a list, when you're working through the paper, of Social Security numbers that you can request a report on. If there's restricted materials and there's a project name or a topic, that can be added to the report that you request from your digital forensic tools.

But it does also mean that reading rooms need to provide tools for the researchers to do this work, and on a very basic level, you need a file viewer tool, because we have just a ridiculous range of file formats, some of them that we don't know what they are because they don't have header information, because they come from a time of computing where there was no header information. You could put whatever extension you wanted on the files. The file extension isn't even going to give you a clue. So there's just a lot of tools that could be used, but if you're not familiar with them, know that even just opening up a file in Notepad, you can access the information. You may not get the way it looked originally, but you can access the information, and that's important.

So, thank you, and I'm looking forward to your questions.

KP: Thank you. (Applause.)

Sarah Conway-Lanz (SC):

Glad to be talking to you this morning. My talk doesn't deal directly with congressional collections, but with an archival processing and digitization project that I hope you'll find relevant to your interests and efforts. Before I came to the Library of Congress, I was an archivist at the Yale Library in the Manuscripts and Archives department there, and had the opportunity starting in 2012 to process a large collection of the Henry Kissinger papers, and prepare them for digitization in a joint Yale/Library of Congress project. It was a two-and-a-half-year project to arrange and describe the papers and prep them for digitization, followed by a yearlong project to digitize and design an online access system for the papers. It was a large collection, at least for Yale, approximately 700 linear feet. I only had a supporting role in the digitization and access system project, but I can try and give you a sense of that work as well. What I'd like to do is give you an overview of the project and some of the goals and the challenges that we had.

So the processing project had some unusual aspects. First, a large portion of the materials were jointly owned by the Library of Congress, so at Yale, we were processing materials owned by another institution, something that's not usually done. It turned out to be a great collaboration. I got to work with Meg McAleer and Alan Ticro in the Manuscript Division here, and build on the work that they'd already done with the Kissinger papers. But because of this joint ownership, it added some unusual work to the process. It wasn't that certain series or segments of the papers were owned by the Library of Congress and others by Yale. It was, for the most part, item by item that the ownership was to be determined. It was self-generated documents, meaning those created by Kissinger and his staff, were owned by Yale, and non-self-generated documents were owned by the Library of Congress, and these item-by-item determinations had to be made before the collection could be arranged and described.

So we spent the first six months of the project determining item ownership. We created guidelines for these decisions and had a team of eight professionals literally writing in pencil "Y" and "LC" on every document. So this is not your usual provenance work with a collection.

This joint ownership had two consequences. It was one of the reasons why the entire collection would be digitized, so that the unified collection could be presented to researchers. But since the digitization and online access portal were being handled by just one institution, Yale, it was also determined that two paper collections with identical content be created. So after digitization, a copy of all the jointly-owned materials were to be printed out, then originals owned by each institution would be interfiled with the copies of the documents owned by the other institution, to create a complete paper copy for the collection for each organization. And this is in addition to sharing the digital files between the two institutions.

So, another unusual aspect at the time, Yale was in the process of establishing a durable and sustainable digital repository and improving the access it provided to the repository for archival collections, especially ones with restrictions to some materials, like the Kissinger papers had. With the Kissinger papers, about 40 percent of the collection was open to all researchers, 60 percent was open with donor permission, and a small percentage had additional restrictions. Yale wanted to be able to provide individualized, online access. This would be based on user ID, with Kissinger and his staff having access to everything online, but other researchers receiving permission at the folder level on a case-by-case basis. Metadata for almost all of the collection would be accessible to everyone, regardless of any restrictions attached to the actual content. Finally, the project had strict deadlines overall. Number one, we had to return LC's materials to them in a timely manner, but also unusually about this project, the project did have the advantage of being well funded, which meant we could hire staff as needed to meet the deadlines.

So once we finished determining item ownership, we started on the normal archival work of arranging and describing the jointly-owned collection. In consultation with LC's Meg McAleer, we established a hybrid arrangement scheme that divided most of the papers into three chronologically-based series – these papers, correspondence, subject files, writings, speeches – into three chronologically-based series – pre-government, government service, post-government. And then those three series were followed by four additional series based on type of material – cartoons, photographs, press clippings, and moving images and sound. Once the physical arrangement was imposed, we wrote up a 500 page-plus finding aid that described the collection down to the folder level. This work took a staff of six archivists and technicians, with the help of about twenty part-time student workers, a year and a half. Then we did similar arrangement description work on a portion solely owned by Yale for the final six months of the project.

But we also did a number of things that manuscripts and archives would have not normally done, because the collection was being prepared for digitization. We did much more work within folders. We alphabetized correspondence within the folders, mainly to ensure that letters were in the correct folder. We removed fasteners to facilitate the scanning. We flagged oversized and unusual materials, like items with Post-it notes attached that we would have to give special instructions to the digital photographers. It's hard to describe the project as MPLP, but we did try to process materials to the minimum level needed.

Digitization affected the arrangement, to a certain degree. For example, we did not combine years of correspondence into a single alphabetical run, something that normally makes serving and using a paper collection easier. Instead, we maintained the arrangement that the papers came in, with alphabetical runs within individual years for decades, for large sections of the materials, and we did this to save processing time.

Arrangement was influenced by restrictions in a few cases, because Dr. Kissinger wanted certain types of documents restricted, and so they had to be placed in restricted sections of the papers. Description was also affected by this digitization prep. We did not break out important correspondence from alphabetical runs. We provided fairly full descriptive notes, but I decided to delay additional description, especially of correspondence, because it could much more easily be done once the collection was digitized. Now, of course, I'm not sure it will ever be done.

We also abandoned the idea of creating the metadata that would have allowed the access system to present the papers as individual, multi-page documents, and as a result, we did not replace fasteners with anything to maintain original associations. So the access system now presents the material a page at a time, first the front, then the back, and even blank backs of documents. So in that sense, the presentation is not any better than that that is afforded in microfilm.

In closing, I wanted to say a few things about the digitization and access system project. The actual digitization work was contracted out to The Crowley Company in Maryland. Yale Library did the quality control. Our preservation department put together an automated tool to facilitate staff work on the quality control. A programmer was also hired to create another automated tool to track workflow for digitization, ingest, and quality control work. The online access system that was created was completely new to Yale. As I understand it, the biggest challenges were adding a context tree to help users. I'll show you the context tree. This is the landing page for the collections, and now we're going into the – of course, it's going to be slow. Oh, and the sizing is not – where is the context tree hiding? It's not even showing on this. Fascinating. Well, there goes my display for you. But the context tree, which helps users to browse and understand collection hierarchies. But the biggest challenge was providing this individualized access to restricted materials.

The system is now working. I promise you, it really is working, with the context tree and all. And the access portal actually takes the requests. It manages the restriction requests and approvals. The portal takes the requests, and approvals are granted through the system on an individual basis with a donor representative, at his home, reviewing the request and making each determination for access. The typed papers from the collections have been OCR'd, and so the collection is full-text searchable. And you can explore the system on your own simply by googling Henry Kissinger papers, and it will come up as the first hit. So, thank you. (Applause.)

Ryan Reft (RR):

Hi, I'm Ryan Reft. I am Sahr's counterpart for domestic policy. I'd be remiss, before I begin, if I didn't just mention that Connie is – all archivists are great, but Connie has processed some of our major collections in law and Congress, and when I have questions, I go to her and the other archivists that really do understand our collections, perhaps better than anyone else. Kathleen, as well, is doing critical work in the digital aspect. I have been in numerous adventures with her in regard to electronic files, which can be circuitous, to say the least. And Sahr and I have worked together as curators on the World War I exhibit and he's really a fantastic colleague. So I would say that it's a buyer's market when they hire here, and so everyone you run into is pretty talented.

We're running out of time, so I'm going to run through this fairly quickly. One, as Karen pointed out, we have over 900 congressional collections. We've added some since 1980, which, in library time, really isn't that long ago. The eighteenth and nineteenth century collections do differ from the twentieth and twenty-first in a number of ways. Without getting too deep into the weeds, they tend to be smaller and more personal, for a number of reasons which I'll get into in a second, whereas the twentieth and twenty-first century collections are much larger for a number of reasons, partially due to the growth of the administrative state with World War I, New Deal, World War II, and the post-war kind of era. By the way, all the pictures on here are from our Prints and Photographs division, so I do encourage you to go up there at some point, and if you have a chance, see the World War I exhibit, because a little soft promotion always helps.

The difference between the nineteenth and twentieth and twentieth and twenty-first is, you know, in the nineteenth and eighteenth congressional collections, you'll have diaries, personal accounts, sometimes family papers. They're smaller; they're more navigable. The twentieth century, because you have the growth of constituent correspondence, patronage files, election campaign ephemera, which on one hand is super interesting from a display/exhibit level, but for researchers sometimes may not be as useful. Subject files on issues, but also case files on state and local concerns. Now, I used to be an urban historian, so I would say that some of the great stuff about the congressional files, whether we're talking about DPM, Daniel Patrick Moynihan, or Patsy Mink, who was on the Honolulu City Council, they are sometimes great for that local kind of – flavor is the wrong word, but legislation and debates – so it also has a local attraction for researchers, as well.

Now, what distinguishes a twentieth century collection? In other words, who do we choose, and why? Well, one, we have moved away over the last thirty or forty years from collecting congressional papers, largely due to the size issue. There may be portions of a collection that are very good, but because of just the sheer arrearage we have and the size of the collection, we really only want certain types of figures at this point. Often, we think that it's better that some of these figures go to their home states, where researchers there can better take advantage of them; they can represent local issues, as well as national issues. That being said, as you see, there's a picture of Edward Brooke. He is significant in part because he was the first African American elected to the Senate since Reconstruction.

So one of the things we look for is first, now, long tenure, pioneering achievement, multiple positions. As Connie pointed out, Jack Kemp, yes, he was a congressman from New York, but he was also HUD security. DPM, Daniel Patrick Moynihan, long-serving senator, but also urban advisor to Nixon, assistant secretary of labor under Kennedy. He wrote speeches for LBJ on occasion, and also ambassador to India. So you're talking about a wealth of experience, and those are the type of figures we really like to veer towards in that regard.

Nancy Pelosi, we don't normally collect Speakers of the House, because sometimes those papers are large, and essentially, what they're doing is herding cats, right, so it's not necessarily policy stuff we're necessarily interested in. But because she's the first female Speaker of the House, we are interested, and she does have – we are processing that collection right now. It is restricted. You have to request access. But nonetheless, it has some great files on San Francisco, particularly when we're talking about BRAC and the demobilization of the military around those areas. Patsy Mink, first woman of color, first

Asian American woman elected to Congress, also would be another one, and her collection is one of our most used, as is Moynihan's.

One other thing to consider besides firsts and these pioneering achievements is that we do look to address areas that, for a number of reasons perhaps, we have not been strong in in the past: Latino and Hispanic Americans, LGBT issues. For example, the last couple years, we accessioned the Frank Kameny papers, who was not in Congress, but nonetheless at the forefront of LGBT issues, both in the homophile and the gay liberation movement, and it is a very juridical collection. So it speaks to our congressional and our judicial collections, which I'll touch on in a second. And, of course, women. You know, attitudes and researching has changed over the last fifty years, thank God, such that we do now pay more attention to figures that in the past were not given the proper research value.

This is just a chart I put together. So I went over our collections and I just looked at the last five years, fifteen collections, and identified our top ones and gave some other ones as kind of outliers. The grey bar is just general usage, and I'll talk about what these things mean in a second. But as you can see, the top five or six are Daniel Patrick Moynihan; Mink; Clare Booth Luce, who I'll talk about in a second; the La Follette family, which was mentioned earlier, Jack Kemp; Emanuel Celler, who is this kind of sleeper collection that I always kind of hype to people because it's really great in a number of ways. He wrote the Hart-Celler Act, was also at the forefront of Israel as a foundation state, and he was in Brooklyn for twenty years. So always interesting, Brooklyn, right? And the Democratic Study Group is another one, which I'll talk about in a second.

So one thing I would couch all those numbers in is that just because we're looking at a congressional figure, people aren't necessarily looking at the congressional files. Like Daniel Patrick Moynihan would be a prime example. And I'd also mention that Daniel Patrick Moynihan, you have to ask permission. That's what kind of makes his numbers amazing, is that he's not an open collection. You have to submit a request and you have to be approved by the donor. Same thing, that was true for Patsy Mink for a period; no longer. But the fact that his use was so high just testifies to the strength of that collection, and Connie is the one who processed that collection.

I would also add that Clare Boothe Luce would be another example. She was only in Congress for two terms, and the congressional papers she has in there are good. But people are really looking at that collection often for Kennedy correspondence, which there is a great deal of, including Joseph when he's over in London during the forties; her art and cultural collections, as well. So it's the full person, and as noted, you see those units, just so you understand what that means, the way our system works is when you just get an analog printout of the usage, it just shows you circulation, the amount of boxes that are requested. By looking at that data, I can't tell you how many different researchers were actually using it for sure. I'd have to go in the analog data for that, which takes more time. But this does give you a sense of just pure circulation and hard numbers, right?

I would say, also, that sometimes numbers can be skewed a little bit. So, like, the Democratic Study Group had a year where it spiked, and I happen to know that year that a certain researcher had a fellowship and was just digging through the boxes for, like, six months. So the numbers can sometimes come off. That's why you have to kind of take the units with a grain of salt.

Now, quickly, there are other ways to get at congressional policy papers, legislation, and these are just some public advocacy groups, like NAACP, which my colleague Adrienne Cannon oversees, which is a fantastic, magnificent collection. People for the American Way, which was mentioned earlier, the Norman Lear organization. It's kind of left of center. That was a new addition. The Democratic Study Group, which I know Julian Zelizer at Princeton has used as a prominent researcher. But then, we have countless advisors, and one thing that's good to mention is that the growth of the presidential library system, with the Wilson administration, we have a number of his cabinet members.

That's less true of later administrations because of the presidential library system. That's not to say we don't have them. It's just, like for example, Don T. Regan, we have the secretary of Treasury's papers, who was also chief of staff. These are just other ways to get at – or Eric Frederick Goldman, who was LBJ's kind of Arthur Schlesinger, right? Perhaps less successful, but nonetheless, that sort of figure. And also, journalists. One of our strongest collections, and I know there are differences in the department about the importance of journalism as a collection, but we acquired David Broeder in 2015, and as Connie was saying, sometimes you get kind of odd objects. Like we have a 1960s cigarette case from when he was covering the JFK – and a 1964 "JFK '64" bumper sticker. Obviously, that didn't happen. So, one thing to consider is that some of these collections have things in them, as Connie said, that are just unusual. Connie is actually now processing the Anthony Lewis collection, which we have an earlier collection that was processed, but we have a much bigger addition. So these are other ways of getting at congressional issues.

Finally, one of our strongest, and actually, collectively, SCOTUS, our collections are used the most, right? We have nearly two dozen twentieth century SCOTUS collections. And this is debatable, but over three dozen – if you go back to the nineteenth century – William J. Brennan, whose correspondence opens this July, for any of you who are interested. We've also added John Paul Stevens, Sandra Day O'Connor, and Ruth Bader Ginsburg in recent years, although all of those – Stevens is closed, and you cannot get access until 2020. O'Connor is closed currently and not taking requests, and Ginsburg you have to request, basically just the ACLU files at the moment, and sometimes she grants them and sometimes she doesn't.

But we also have Robert Bork, Harold Leventhal, Skelly Wright, all circuit court judges for the Columbia District, and then other prominent legal figures like Joseph Rauh, who, though not a congressional figure, is a leading civil rights lawyer, head of the ADA. So those are just other ways. And, you know, one thing to consider about SCOTUS is there is a line of jurisprudence that says we should look at what the congressional debates were behind legislation to understand what the meaning and the motive, and obviously, originalists disagree with that. But nonetheless, that is another way of looking at it, because some of the justices do pay close attention to those issues.

In the interest of time, I'm going to kind of just quickly – we've already talked about Daniel Patrick Moynihan. Patsy Mink, I think one of the strengths was her Title IX work, gender equity, but also education, welfare, and those things. And finally, in our newest additions, and these are all restricted and you can request access – and one thing about the organizations, too, it was mentioned earlier to look at advocacy groups. Well, you know, we have William Baroody's papers, the founder of AEI, American Enterprise Institute, and we actually are in the process of processing Edwin Feulner, the former, and now, apparently, current, – head of the Heritage Foundation, and one of the founders.

Nancy Pelosi's papers, we are processing. We believe they'll be done in mid-June. Those also are closed and will require a permission request. And just as an aside, in terms of advisors, we also have John A. Lawrence, who was her chief of staff in 2008 to 2013. Now, that's unprocessed, and it also requires a permission request. But again, these are the directions we're moving in in terms of future research, and things you might want to consider as you think about research down the line.

And that is – that's it. (Applause.)

KP: Thank you, panelists, for excellent presentations, and I think we have time for a couple of questions. Yes?

US: So, do you actively pursue people's work, papers, or do you wait for them to come to you?

RR: Is that a question for me?

US: Anyone who wants to answer. And who's on your wish list?

RR: Am I allowed to say who's on our wish list, Janice? No. That's why it's a wish list, because it's in my mind. And Sahr can comment on this, too, because Sahr does, also, acquisition papers. We do have people in mind. We have reached out to certain individuals that I won't say at this moment. But we absolutely do pursue new collections, although at the moment, we are focusing on trying to process. We have a lot of arrearage and a lot of great collections that we need to process. But, yes, we do send out inquiries. Do you want to add to that?

SC: Well, since I don't do congressional collections, I'll let you –

RR: No, but you do acquisitions.

SC: Yeah, no, don't have anything to add.

KP: Any other questions? Yes.

US: Do you ever have opportunities when your electronic content affects your appraisal decision for whether or not to keep paper? Do you talk to each other about (indiscernible 00:57:59)?

CC: Yes, we do. I think that actually happened more with the Jack Kemp papers, and that was before Kathleen was working, and I was working with Karen Stewart, our automation person at the time. But yes, sometimes that does come up, and we discuss that, yes.

US: And does that happen also with media that you discover elsewhere, like (indiscernible 00:58:28) recordings that might be duplicated within your collection if they're available somewhere else?

CC: We have done that, yes. And sometimes it depends on what the instrument of gift says, because some of the duplicates we did offer back to the Kemp family, and they took some of the materials themselves. So we offered it back to them, and they took it.

KO: And they're also transferred within the library, depending on the format. So there will be, sometimes, serials that we get that complete a serial in another part of the library, and with the American Lands Alliance, that was one of the first times we transferred digital geospatial material to Geography and Maps.

KP: Anyone else?

US: How difficult was it to negotiate with Yale to build a partnership?

SC: So I came on the project after all of that hard negotiation was done. Certainly, my impression was it was complicated to get two institutions to work together in this way, but it ended up working out smoothly in the end. So, thankfully, that hard work was done before I started the project.

US: Do you have a preservation plan for continued accessibility of your digital collection?

KO: I don't think we have a formal plan. Right now, we're committed to bit-level preservation of the materials, because that's what we think we can do responsibly and efficiently and with the resources that we have. I should say, from my own perspective, I think that the tools coming up, a lot of which are open source, including emulation and file viewer tools, I think will make a lot of these materials accessible long term. Of course, it does depend on the format. And really, with the file viewer tools, there's been very few that we haven't been able to access. And generally, when we haven't been able to access it, it usually isn't a file format question, it's a question of the media itself being corrupt.

But we do have long-term storage. We have them saved in two different locations on two different systems. We have checksums that we generate at the very beginning of our process, and that fixity is checked the entire time it's on. But in terms of access preservation, I think that's still the Wild West for a lot of institutions.

US: I have just another quick question that touches on something that you mentioned in passing, so I was just curious about a little bit more detail. And it has to do with (indiscernible 01:01:37) access, and not so much to do with the file (indiscernible 01:01:40) tools, but you mentioned sort of (indiscernible 01:01:42) the content in bags, and I wonder if it's actually happening. So they request a bag of material, or is it (indiscernible 01:01:52) or how does that actually work?

KO: So, again, in terms of efficiency, we decided we weren't going to allow specific file requests. It was going to be served as we saved it.

US: So that is the bag.

KO: It is the bags, yeah.

KP: Anyone else? Well, I think our time is more than up, and thank you very, very much.

[End of interview]