Digitizing History:

Collaborating to Digitize the United States Congressional Serial Set

LAW LIBRARY
LIBRARY OF CONGRESS

GPO

Speakers

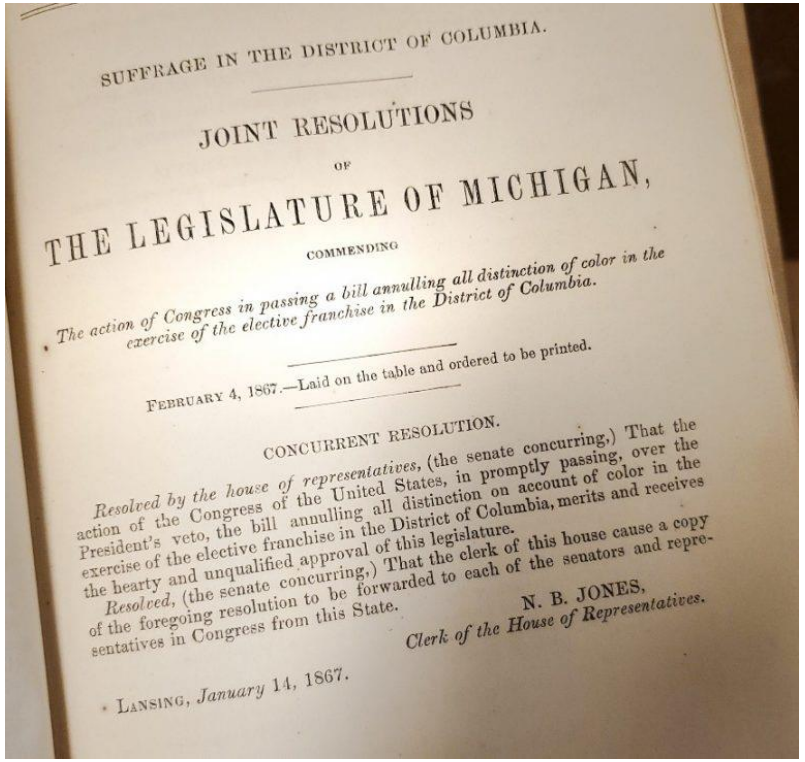
- James Sweany, Chief, Digital Resources Division, Law Library of Congress
- Chris Ehrman, Digital Project Coordinator, Law Library of Congress
- Bailey DeSimone, Library Technician (Metadata), Law Library of Congress
- Elina Lee, Library Technician (Metadata), Law Library of Congress
- Heidi Ramos, Lead Program Planner, U.S. Government Publishing Office
- Suzanne Ebanues, Supervisory Management Analyst, U.S. Government Publishing Office

Law Library Serial Set Team

- Introduction - James Sweany
- Evaluation and Workflows – Chris Ehrman
- Serial Set Database – Elina Lee
- Documentation and Promotional Activities – Bailey DeSimone

Digitization Status of the U.S. Serial Set

- LAW & GPO have started a large, 10-year effort to digitize and make accessible volumes of the US Congressional Serial Set back to the first volume, published in 1817.
- Pilot Project: 1,069 volumes digitized during Fiscal 2019.
- Plan is to digitize an estimated 15,850 volumes spanning 177 years and including 73,000 maps and foldouts.



SUFFRAGE IN THE DISTRICT OF COLUMBIA.

JOINT RESOLUTIONS

OF

THE LEGISLATURE OF MICHIGAN,

COMMENDING

The action of Congress in passing a bill annulling all distinction of color in the exercise of the elective franchise in the District of Columbia.

FEBRUARY 4, 1867.—Laid on the table and ordered to be printed.

CONCURRENT RESOLUTION.

Resolved by the house of representatives, (the senate concurring,) That the action of the Congress of the United States, in promptly passing, over the President's veto, the bill annulling all distinction on account of color in the exercise of the elective franchise in the District of Columbia, merits and receives the hearty and unqualified approval of this legislature.

Resolved, (the senate concurring,) That the clerk of this house cause a copy of the foregoing resolution to be forwarded to each of the senators and representatives in Congress from this State.

N. B. JONES,
Clerk of the House of Representatives.

LANSING, *January* 14, 1867.

(H. Misc. Doc. 42, 39th Cong., 2d Sess., at 1 (1867) reprinted in Serial Set vol. 1302). Photo by Bailey DeSimone.

Serial Set Partnership – GPO and the Law Library

- **Law Library:**
  - Manages the digitization of Serial Set volumes
  - Provides accepted digitized volumes to the GPO

- **Government Publishing Office (GPO):**
  - Will provide cataloging of the Serial Set volumes and documents
  - Adding authentication to the PDF files
  - Providing Serial Set volumes that are missing or incomplete in the Library's collections

- The digitized Serial Set will be available through both the Library of Congress and GPO websites (**govinfo**).

Digitization Status of the U.S. Serial Set

Performance Target: Expand the digitization of law collections to better meet the demand for online access to collections by continuing the digitization initiative, and digitizing 3,500 volumes of the U.S. Serial Set by September 30, 2020.

Quarter Milestone	Accomplishment
Q1: Prepare 875 volumes and send to digitization	1,094 volumes had been prepared for shipment to the vendor.
Q2: Prepare 875 volumes and send to digitization (1750 total)	A total of 2,351 volumes had been prepared for shipment to the vendor
Q3: Prepare 875 volumes and send to digitization (2625 total)	A total of 2,351 volumes had been prepared for shipment to the vendor.
Q4: Prepare 875 volumes and send to digitization (3,500 total)	A total of 2,602 volumes had been prepared for shipment to the vendor.

Digitization Status of the U.S. Serial Set

FY21 Goals:

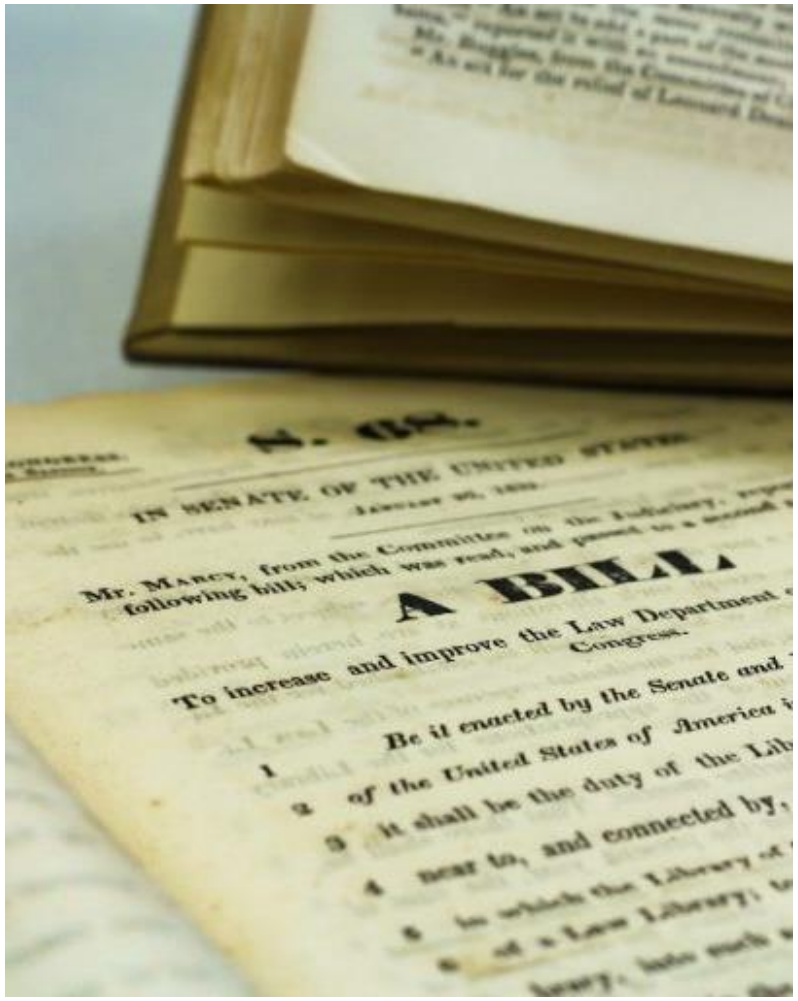
- Prepare and digitize 3,000 volumes
- Publishing a subset of the Serial Set (69th Congress) as a P1 digital collection.

US Serial Set Physical Characteristics

- Digitizing 15th through the 103rd Congresses
- Over 15,700 volumes containing around 12 million pages
- Range of page count of a couple dozen pages to around 2000 pages
  - Foldouts
  - Maps
  - Oversized maps

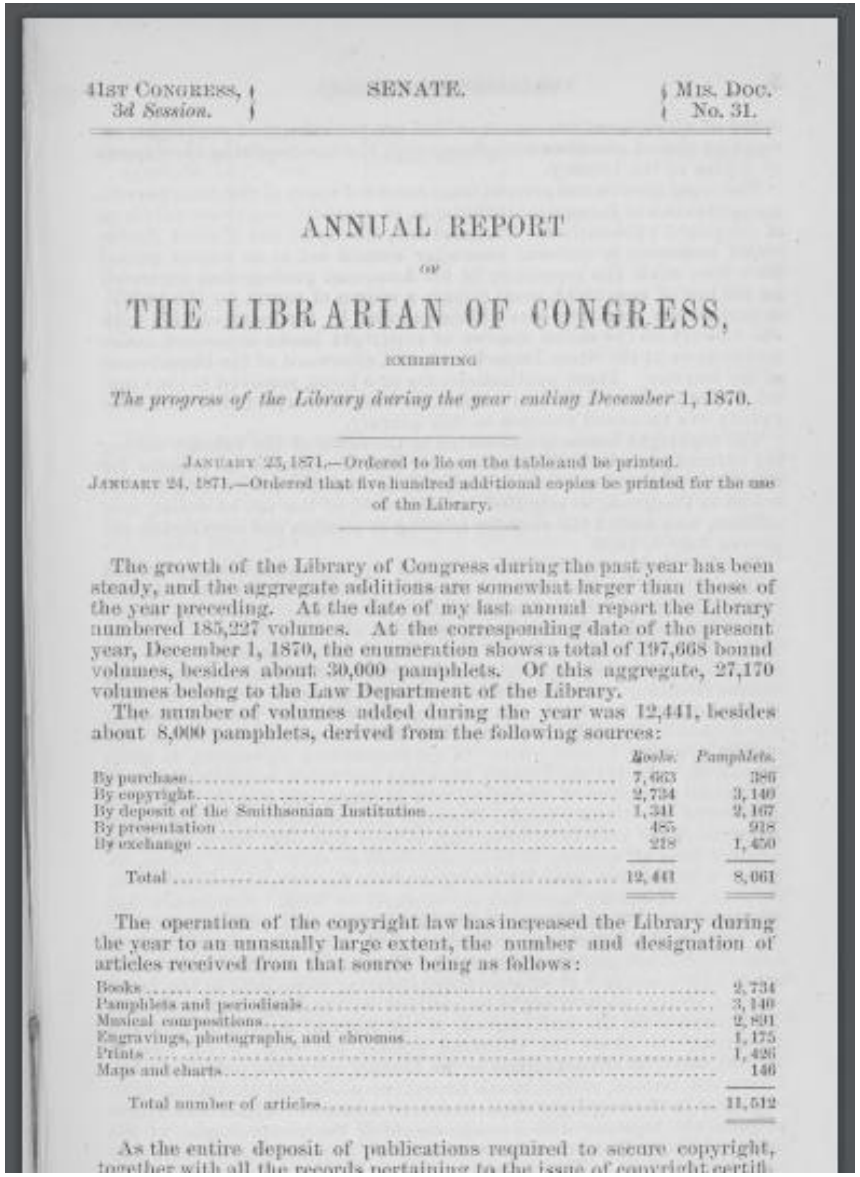
Volume Evaluation for Digitization

- Up to six copies in the Library of Congress collections
- Inspect physical volumes for completeness
  - Expected documents from 1909 Index and Schedule of Documents
- Record physical characteristics
  - Foldouts
  - Maps
  - Color plates
  - Oversize foldouts
- Request volumes stored offsite
- Request volumes from GPO
- Create complete Serial Set number from multiple volumes



Digitization Specifications

- FADGI 3-Star performance level
- Digital objects
  - Master file: JPEG2000 (JP2)
  - OCR: XML
  - Thumbnail: GIF
  - Document: PDF/A 2b
  - Volume: PDF/A 2b
- Unique bag id
  - Serial Set Number_Part_Volume
  - 08858_6f_00
  - Volume PDF name
- Document directories
  - Sequence-document number-part
  - 08739_01_00-005-0557-0000/
  - Document PDF name
- Delivered in BagIt structure



41ST CONGRESS, 3d Session. | SENATE. | MIS. DOC. No. 31.

ANNUAL REPORT

OF

THE LIBRARIAN OF CONGRESS,

EXHIBITING

The progress of the Library during the year ending December 1, 1870.

JANUARY 25, 1871.—Ordered to lie on the table and be printed.
JANUARY 24, 1871.—Ordered that five hundred additional copies be printed for the use of the Library.

The growth of the Library of Congress during the past year has been steady, and the aggregate additions are somewhat larger than those of the year preceding. At the date of my last annual report the Library numbered 185,227 volumes. At the corresponding date of the present year, December 1, 1870, the enumeration shows a total of 197,668 bound volumes, besides about 30,000 pamphlets. Of this aggregate, 27,170 volumes belong to the Law Department of the Library.

The number of volumes added during the year was 12,441, besides about 8,000 pamphlets, derived from the following sources:

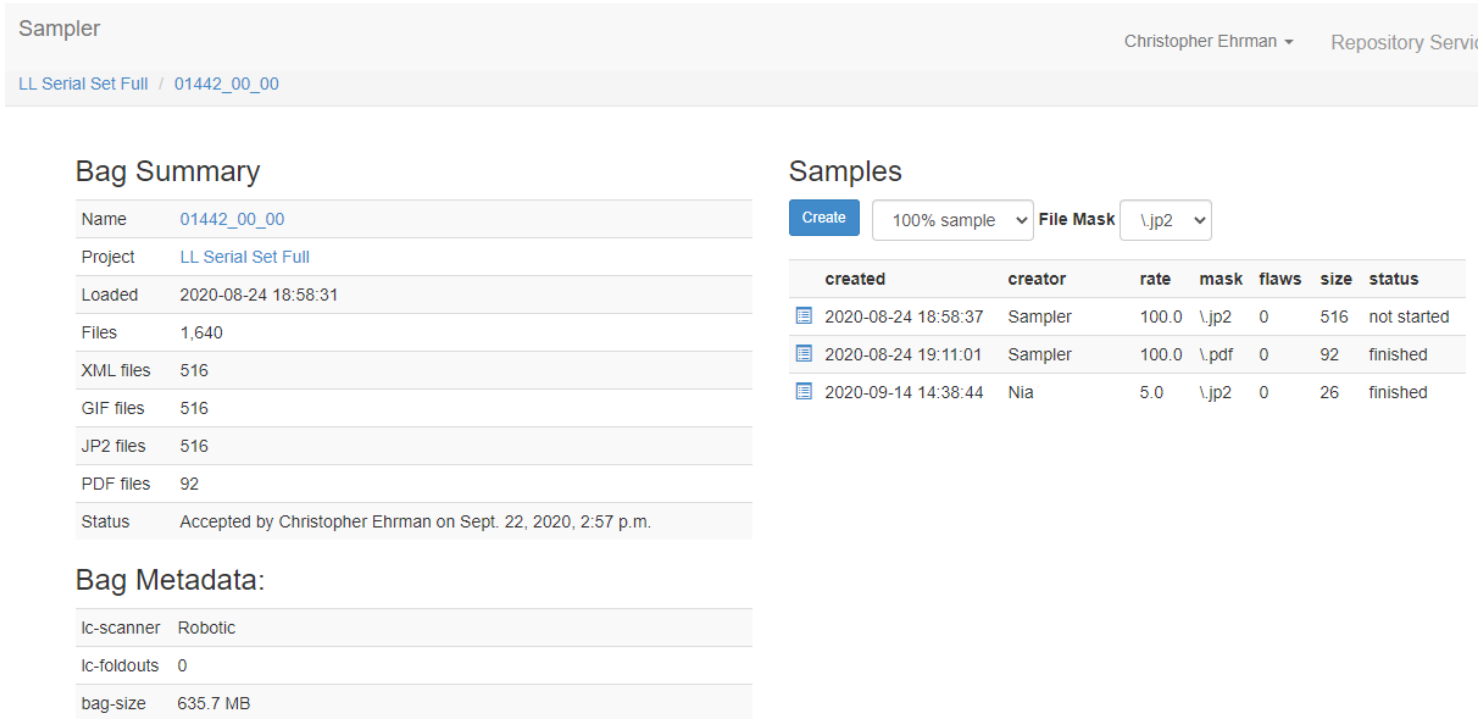
	Books.	*Pamphlets.*
By purchase	7,663	386
By copyright	2,734	3,140
By deposit of the Smithsonian Institution	1,341	2,167
By presentation	485	918
By exchange	218	1,450
Total	12,441	8,061

The operation of the copyright law has increased the Library during the year to an unusually large extent, the number and designation of articles received from that source being as follows:

Books	2,734
Pamphlets and periodicals	3,140
Musical compositions	2,891
Engravings, photographs, and chromos	1,175
Prints	1,426
Maps and charts	146
Total number of articles	11,512

As the entire deposit of publications required to secure copyright, together with all the records pertaining to the issue of copyright certifi-

Digitized volume evaluation

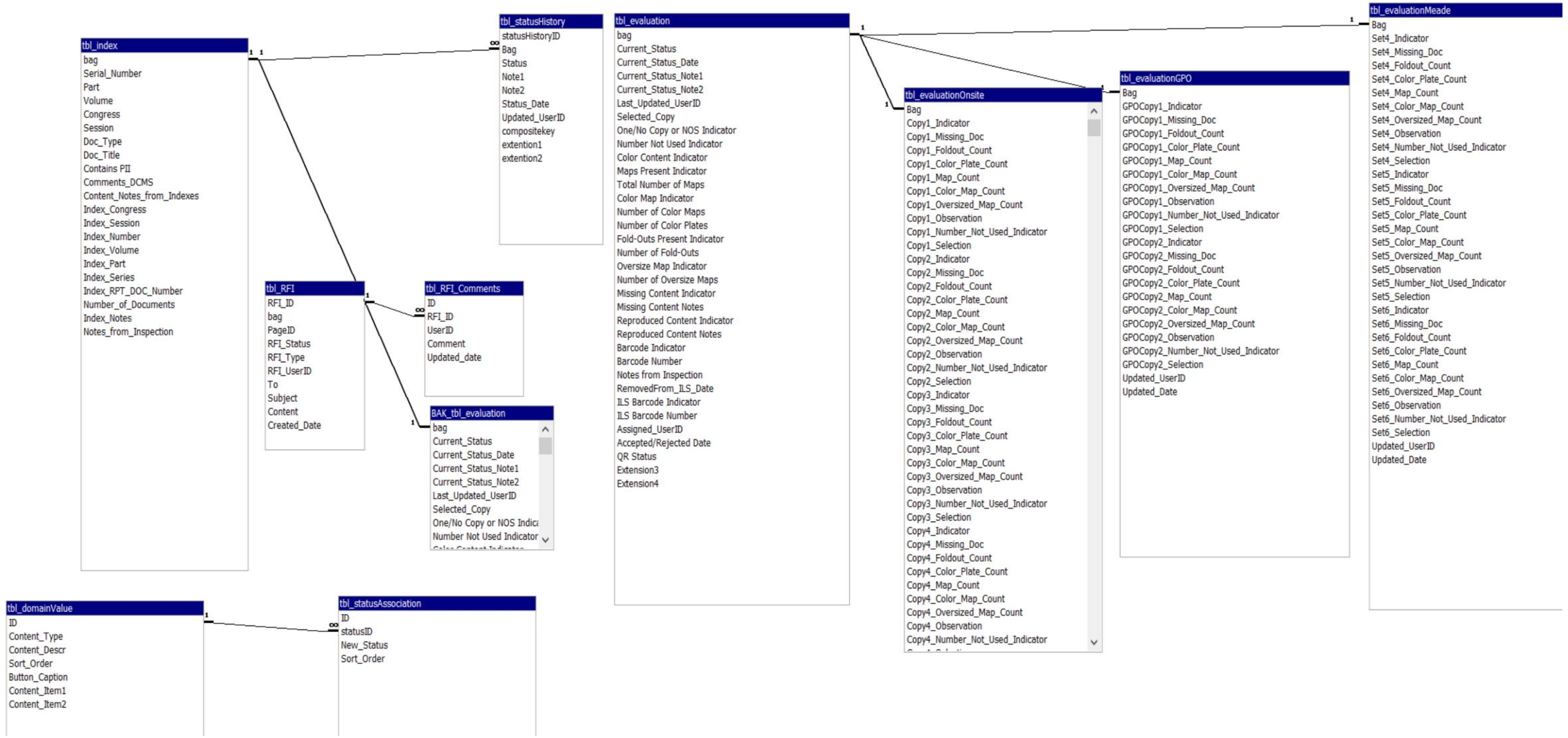


- Ingest into the Library's digital content management system
- Quality review in Sampler
  - Images, directory structure, completeness

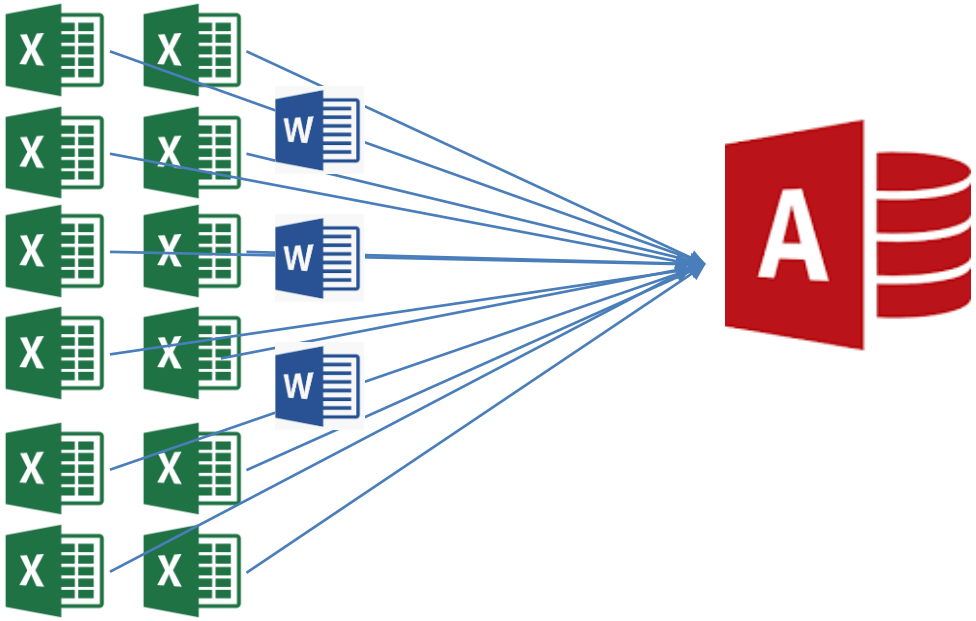
Next steps

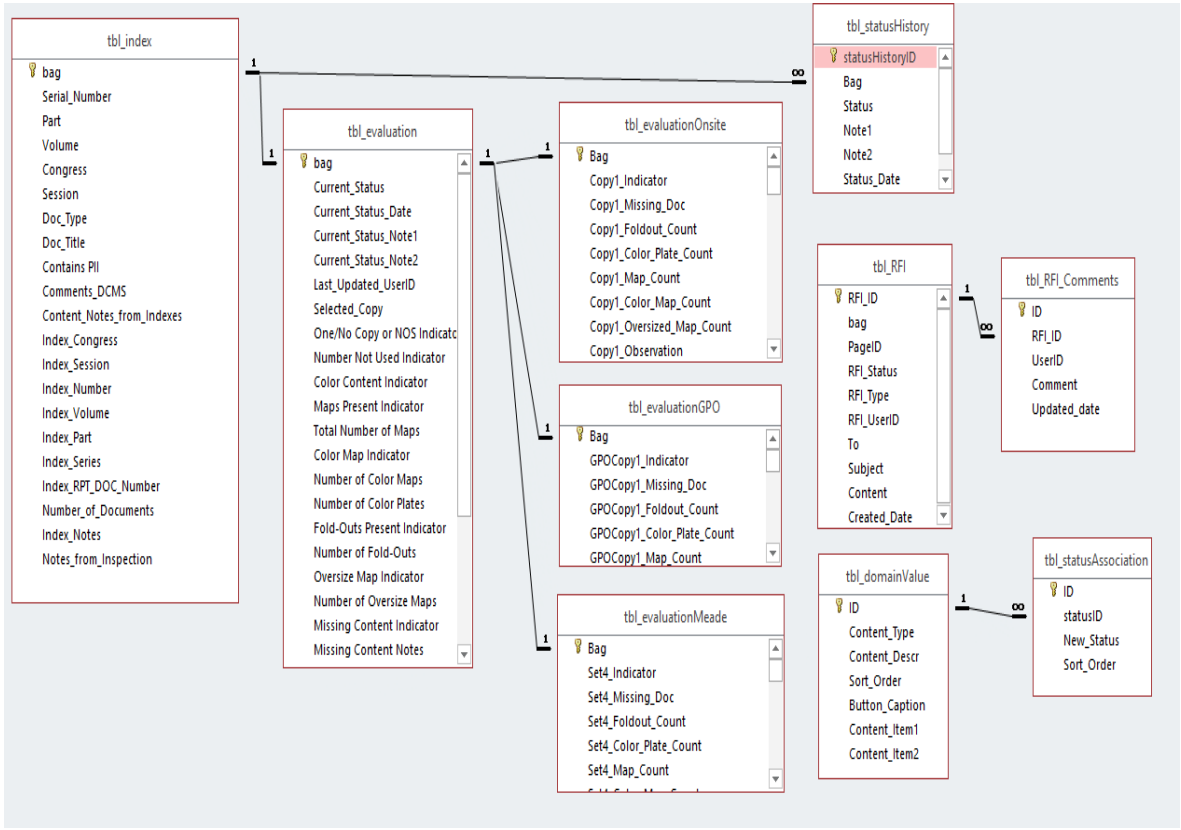
- Send accepted bags to GPO
- GPO
  - Create MARC records
  - Authenticate PDF files
  - Send to Library of Congress
- Ingest authenticated PDFs into the digital content management system
- Presentation in loc.gov

ERD (Entity Relationship Diagram)



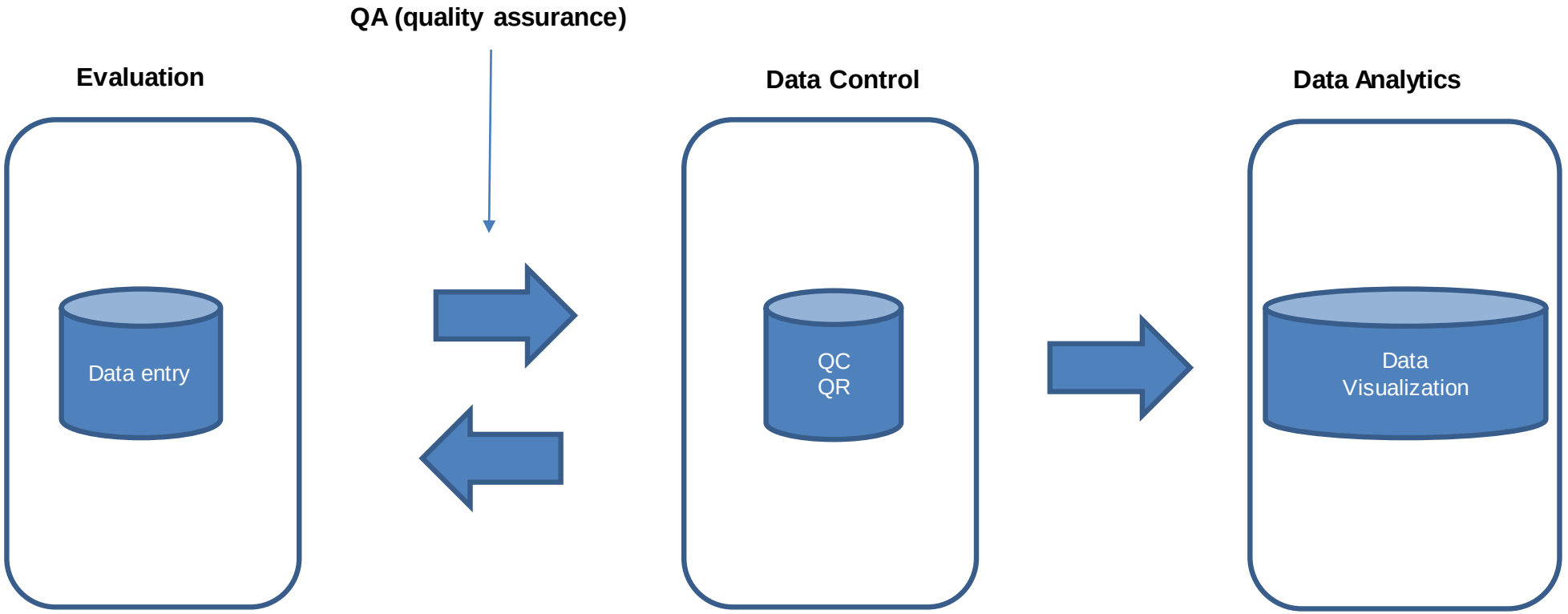
Data Centralization





- Consolidate 33 Excel files, 3 Word files, 17 Reports
- Up to 200 multiple users enter data simultaneously
- Real time update
- Tracking all information history/Reduce human errors

Workflow



Onsite Evaluation
Ft. Meade Evaluation
GPO Evaluation
Digitized Volume Evaluation

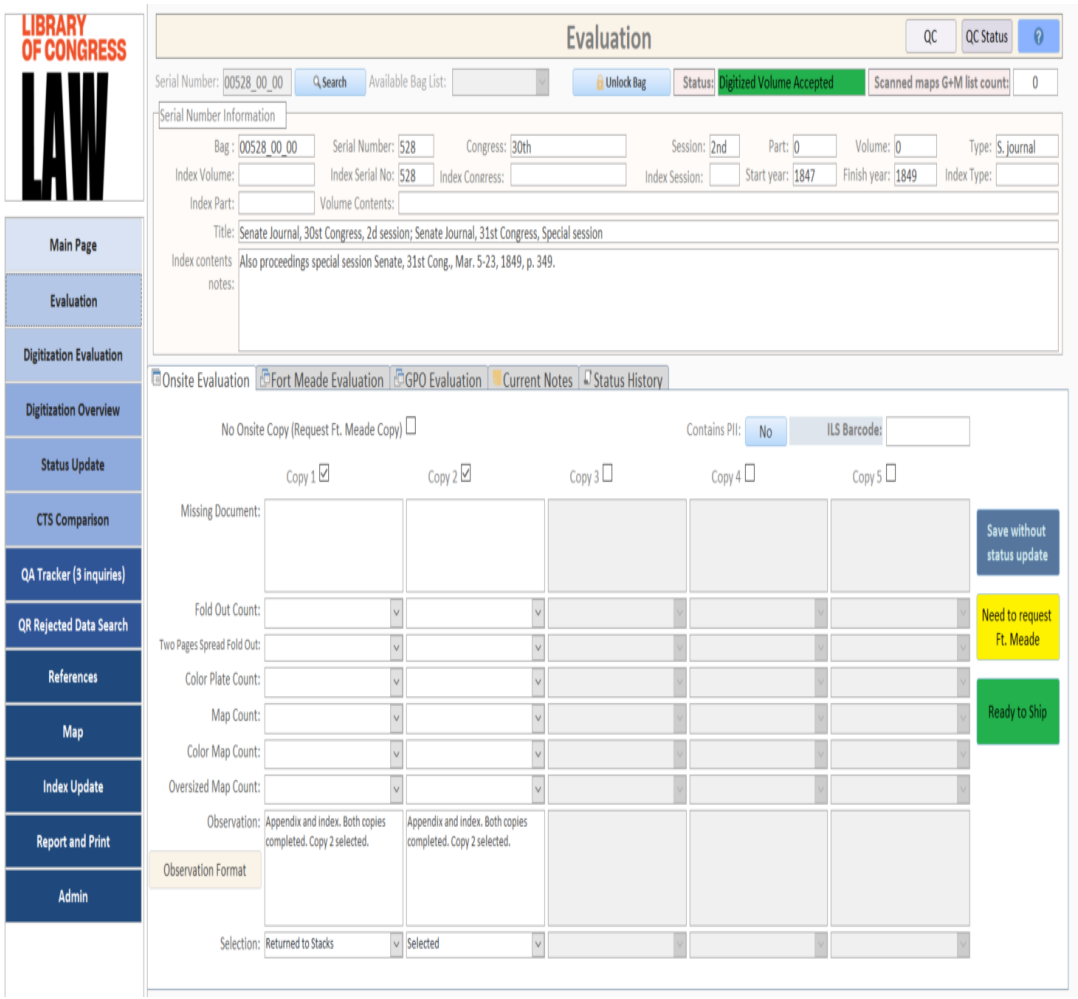
Quality Control/Quality Control Resume
Data analysis
Data discovery/error resolution

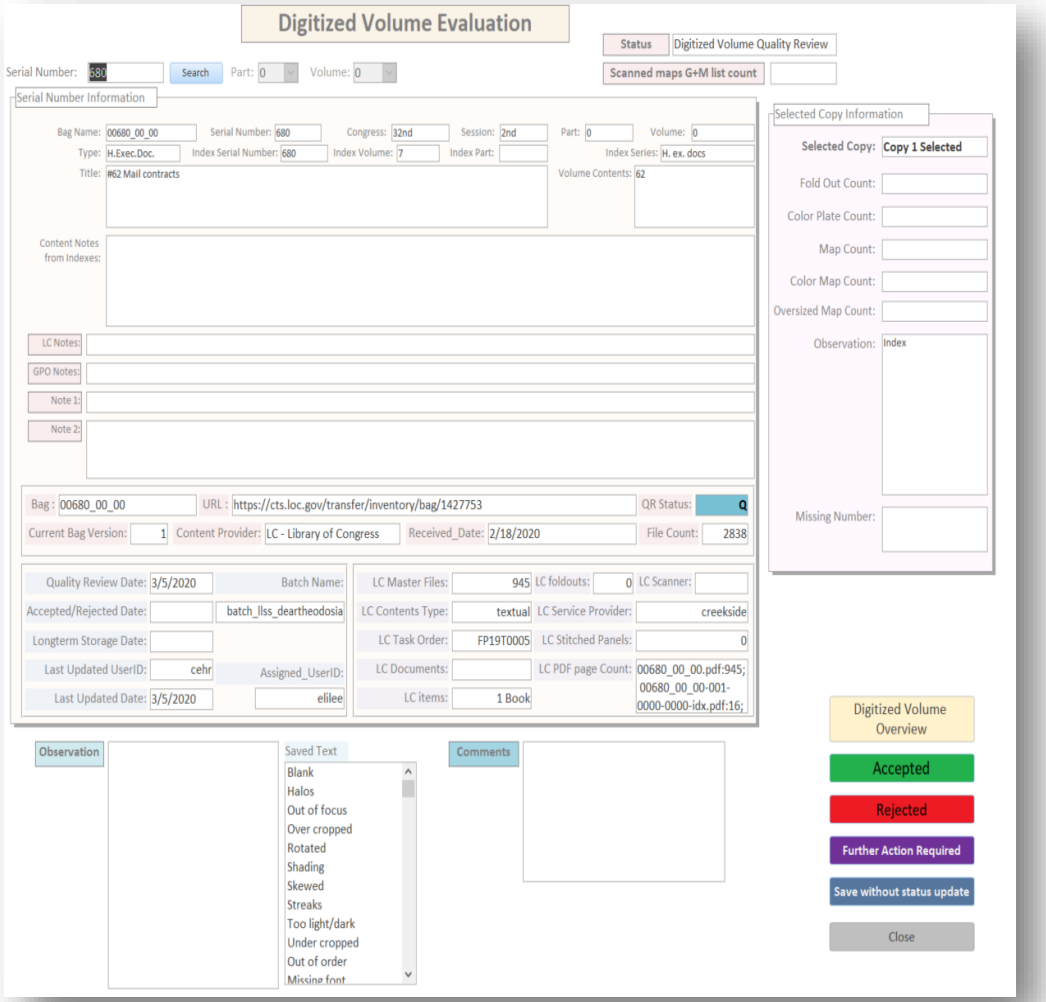
Tableau
DRD Tableau Public

Serial Set Database V.1.0 (Aug. 2019)

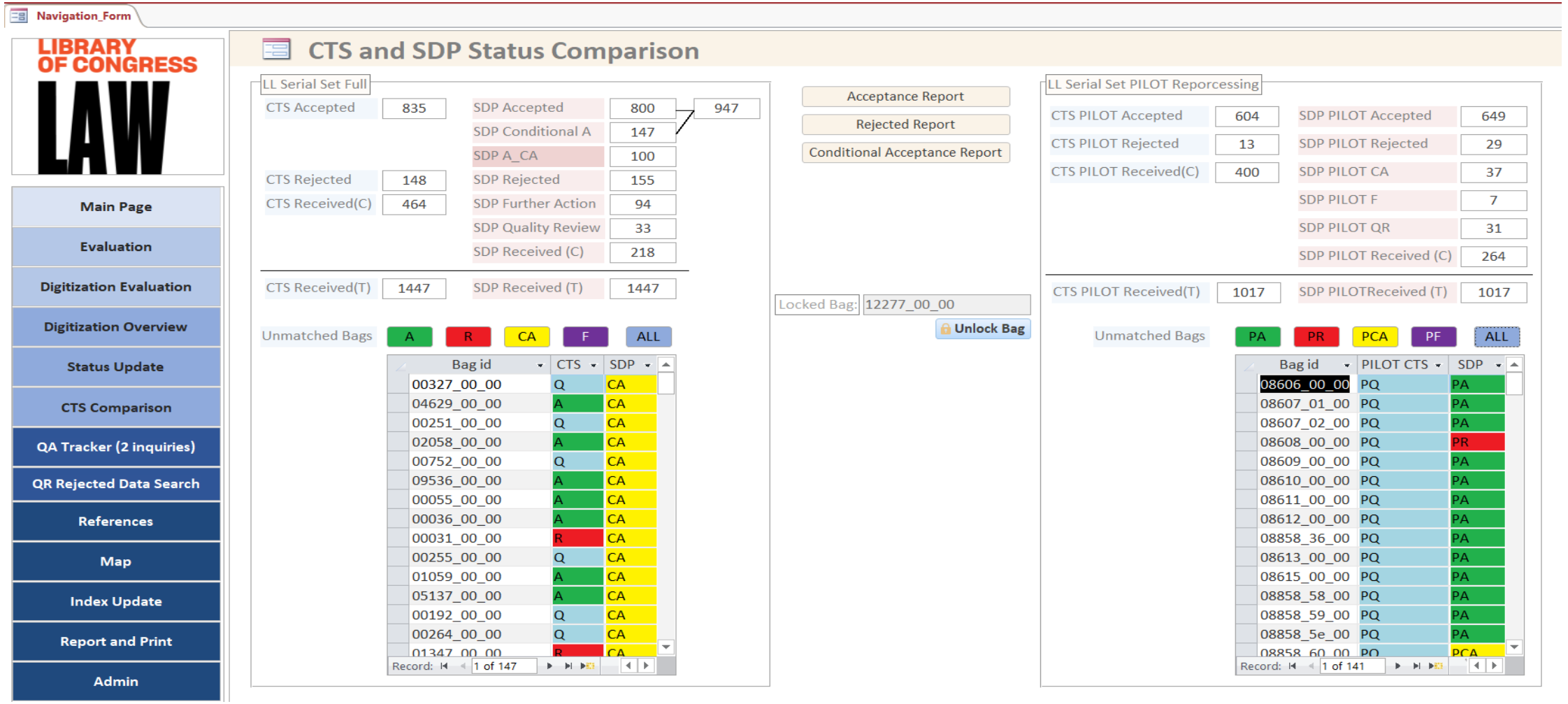
Serial Set Database V.2.0 (Aug. 2020)

Evaluation Entry Form





Data Monitoring



Print Sheet and Report

Shipment Report

Sunday, August 9, 2020

Shipment: 3 Box_Number: 203

Bag: 01505_00_00 **Shipped date: 8/6/2020** **Serial Number: 1505** **Part: 0** **Volume: 0**

Congress: 42nd | Session: 2nd | Doc Type: H.Exec.Doc. | Maps: | Observation | Note
Index Volume: 3 | Index Part: 1 | Vol. Contents: 1, pt. 5 | Color Maps: | ID:2593
Doc Title: Annual report of Secretary of Interior, 1871, and reports of subordinate officers | Color Plates: | Index
Fold-Outs: 2 | Loose pages
Book tracings
Better fold outs
Index Notes: Interior rp., 1871, v. 1. | Oversize Maps:

Bag: 01507_00_00 **Shipped date: 8/6/2020** **Serial Number: 1507** **Part: 0** **Volume: 0**

Congress: 42nd | Session: 2nd | Doc Type: H.Exec.Doc. | Maps: 2 | Observation | Note
Index Volume: 4 | Index Part: | Vol. Contents: 2; 1, pt. 3; 1, pt | Color Maps: | ID: 2597
Doc Title: #2 Annual report of Secretary of Treasury on State of finances, 1871 #1/6 Annual report of Secretary of Navy, 1871 [following document numbered 2] #1/7 Annual report of Postmaster General, 1871 [following second p. 318] | Color Plates: | Index
Fold-Outs: 2 | Dark pages
Book tracings
Better foldouts
Index Notes: Navy, 1871; Post-Office, 1871. | Oversize Maps:

Bag: 01508_00_00 **Shipped date: 8/6/2020** **Serial Number: 1508** **Part: 0** **Volume: 0**

Congress: 42nd | Session: 2nd | Doc Type: H.Exec.Doc. | Maps: | Observation | Note
Index Volume: 5 | Index Part: | Vol. Contents: 3, 4 | Color Maps: | ID: 2598
Doc Title: #3 Annual report of Comptroller of Currency, 1871 #4 Annual report of Commissioner of Internal Revenue, 1871 | Color Plates: | Index
Fold-Outs: | Brittle, torn pages
Loose pages
Ink spots
Index Notes: | Oversize Maps: | Complete

Bag: 01509_00_00 **Shipped date: 8/6/2020** **Serial Number: 1509** **Part: 0** **Volume: 0**

Congress: 42nd | Session: 2nd | Doc Type: H.Exec.Doc. | Maps: | Observation | Note
Index Volume: 6 | Index Part: | Vol. Contents: 5-20 | Color Maps: | ID: 2600
Doc Title: House Executive Documents | Color Plates: | Index
Fold-Outs: | Torn, brittle pages
Stains on pages
Index Notes: | Oversize Maps: | Ink spots
Complete

Bag: 01510_00_00 **Shipped date: 8/6/2020** **Serial Number: 1510** **Part: 0** **Volume: 0**

Congress: 42nd | Session: 2nd | Doc Type: H.Exec.Doc. | Maps: | Observation | Note
Index Volume: 7 | Index Part: | Vol. Contents: 21-84 | Color Maps: | ID: 2602

Statistics and Data Analytics

- DRD Tableau Public
  Manages and tracks whole Serial Set
- Tracks Quality Control
- Monthly, quarterly, yearly progress reports and projections

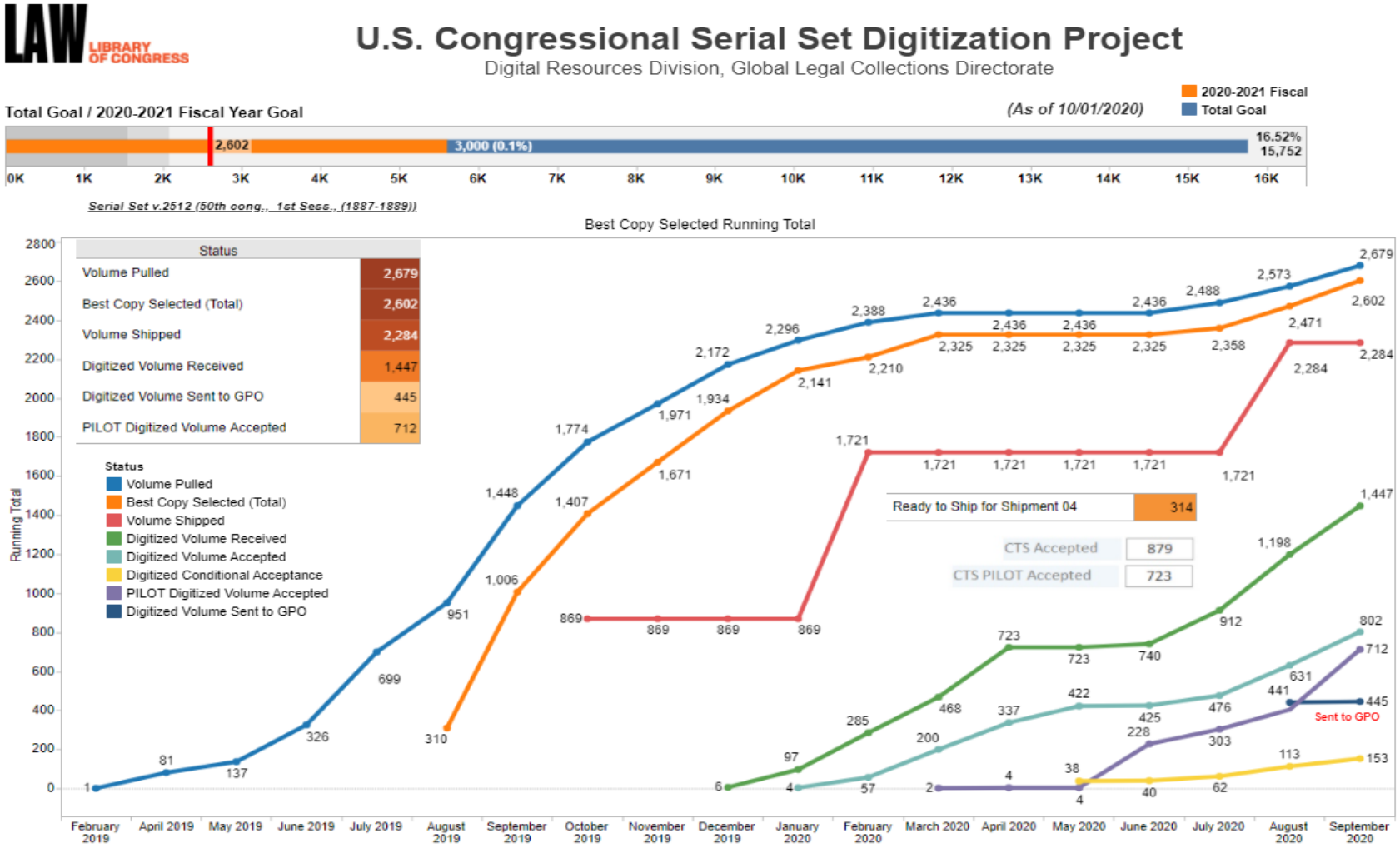
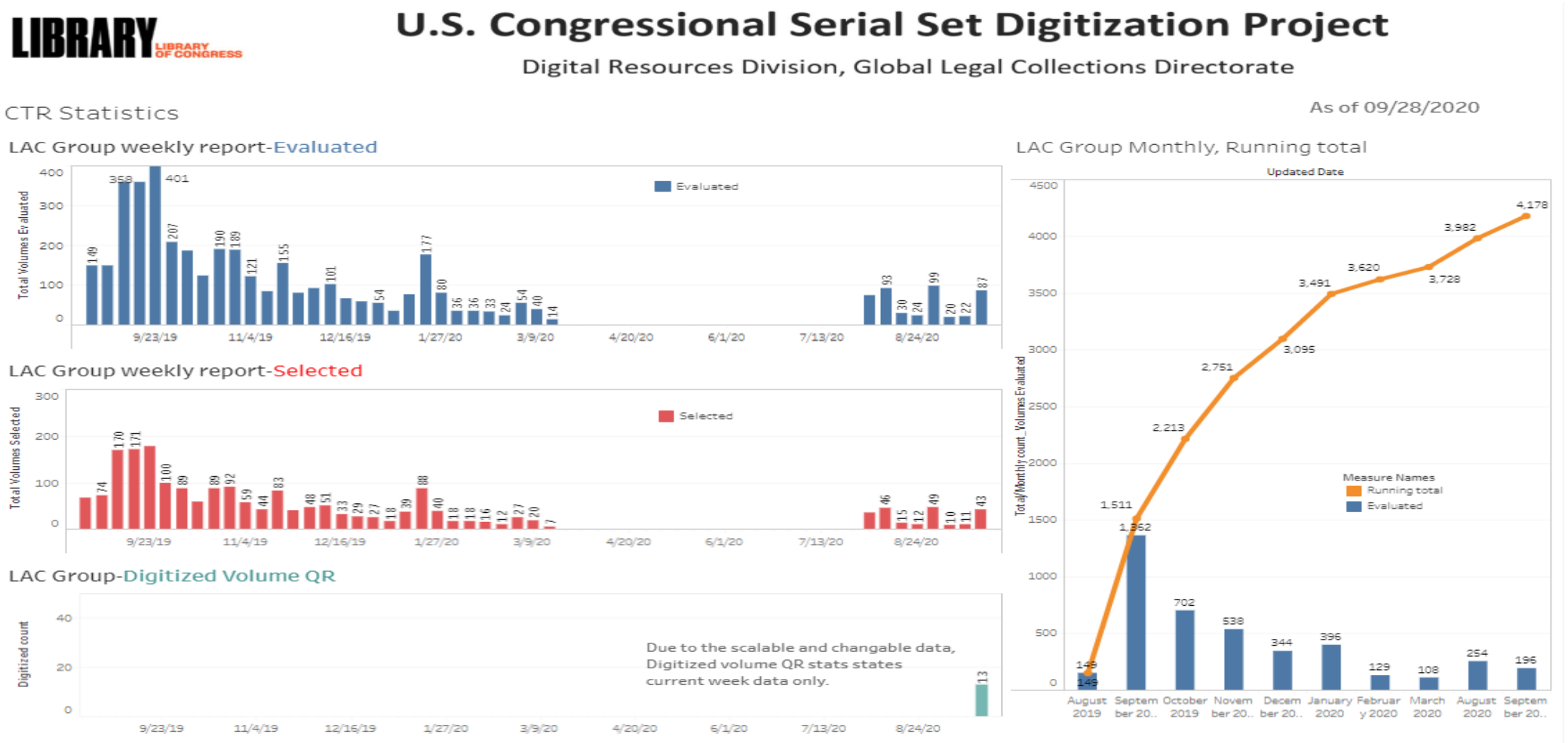
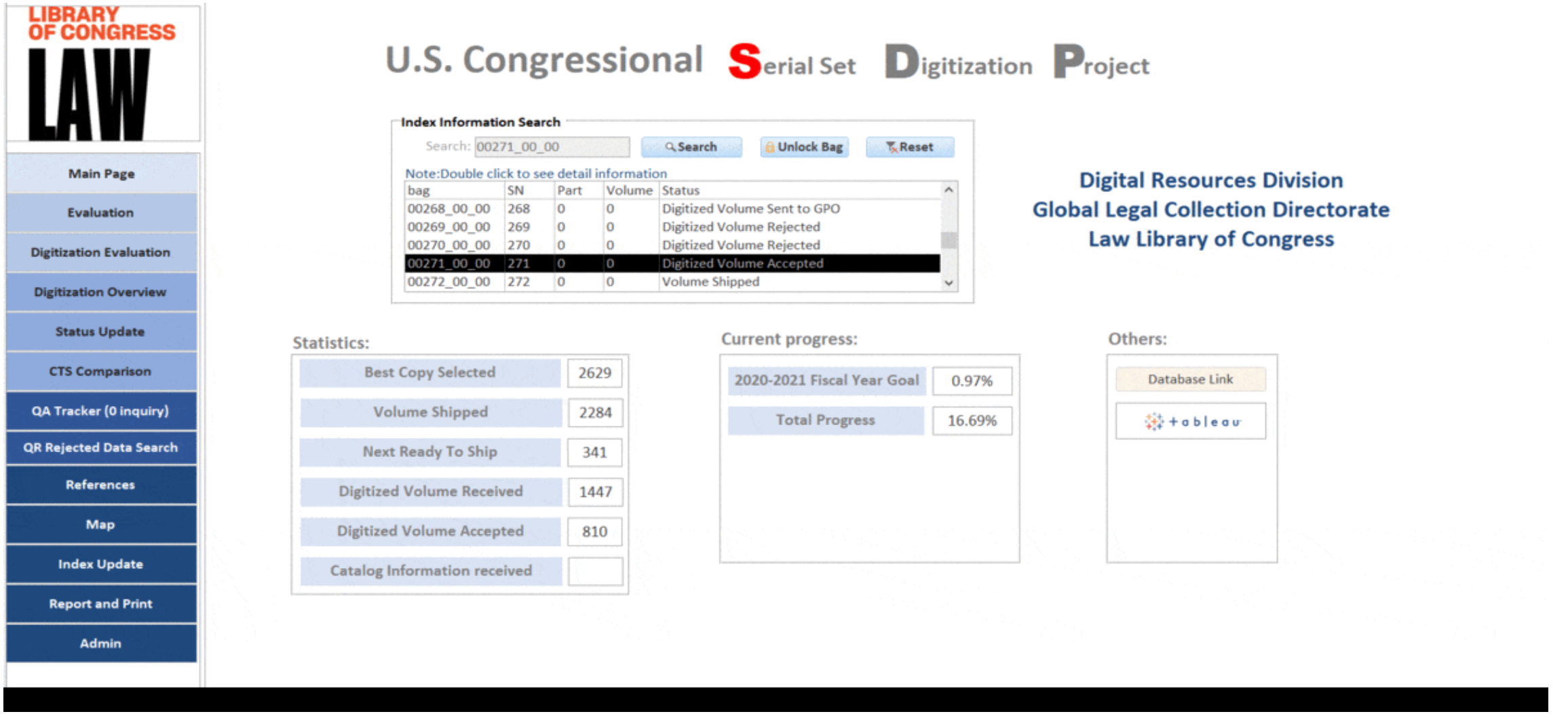


Tableau: Contractor Statistical Report



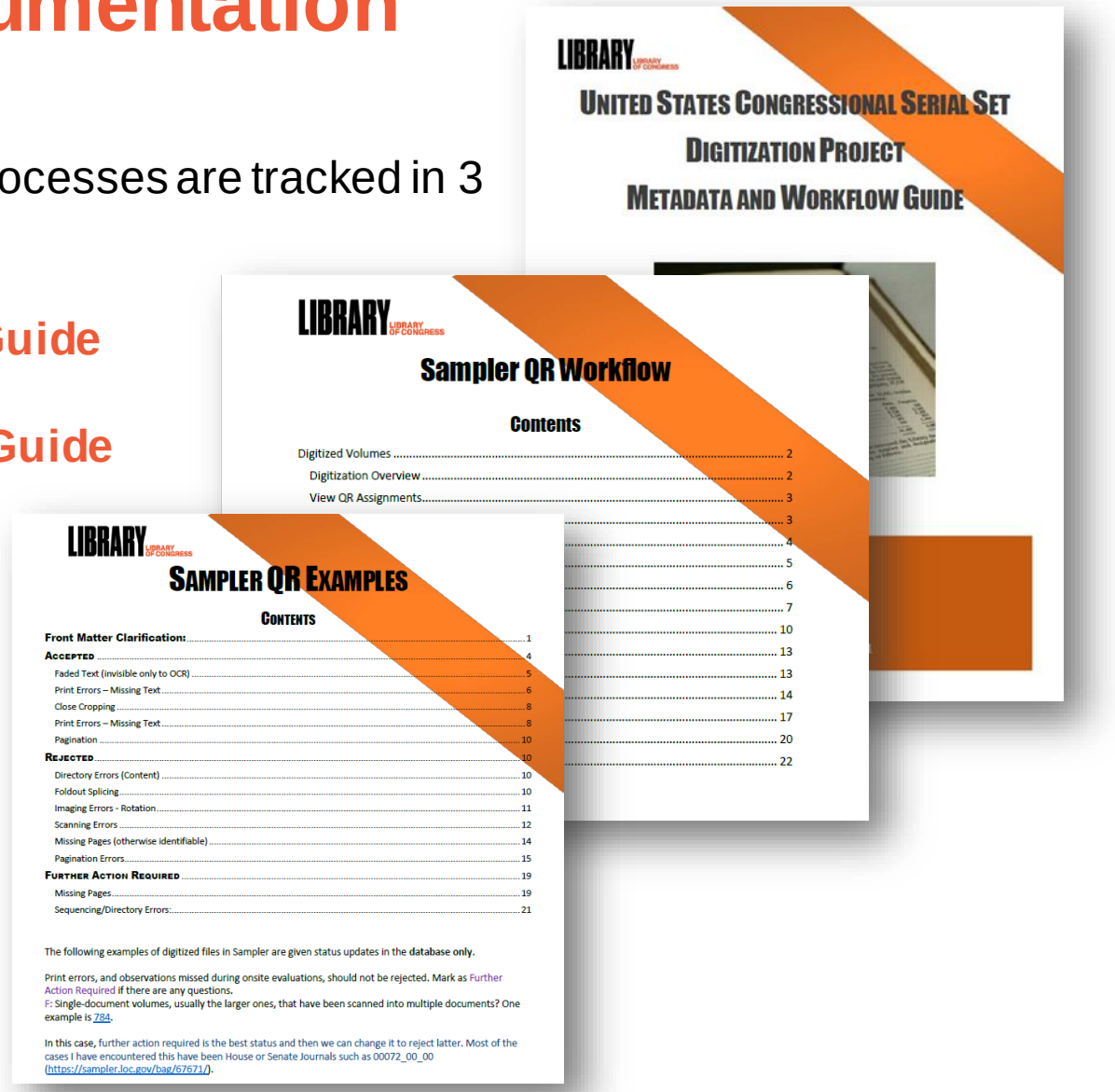
Serial Set Database Overview



Workflow Documentation

Instructions for workflow processes are tracked in 3 guides:

- **Serial Set Digitization Guide**
- **Sampler QR Workflow Guide**
- **Sampler QR Examples**

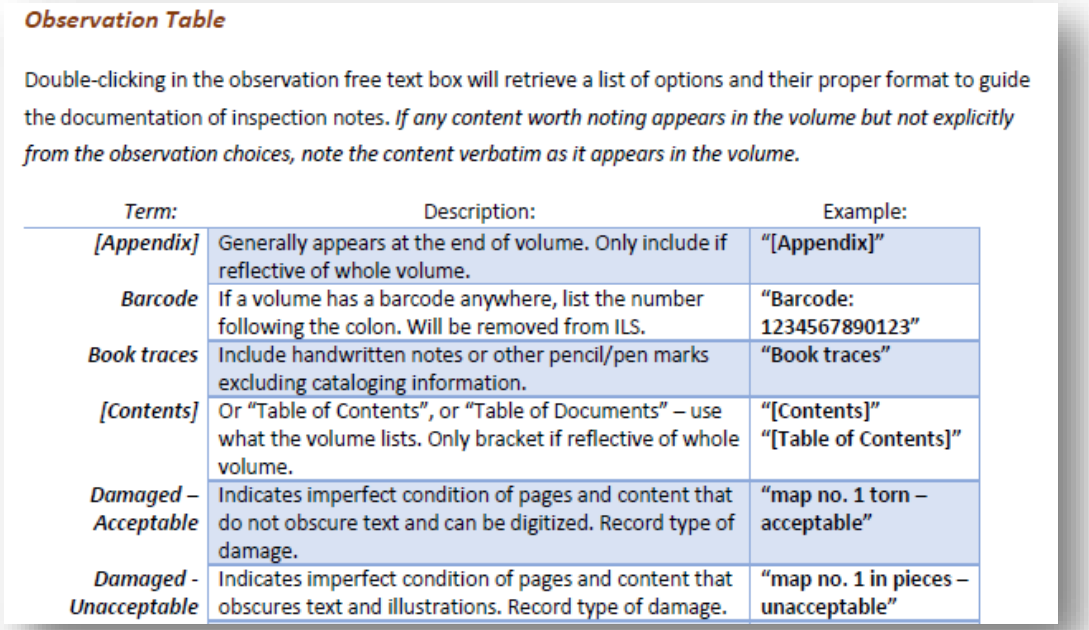


Metadata Guide: Identifying

Instructions for documenting and formatting metadata are included in the guides.

These tables are updated regularly as new information that should be recorded is discovered in the Serial Set.

The goal of metadata tracking is to create unique terms for Serial Set content so that it can easily be retrieved through a query search.



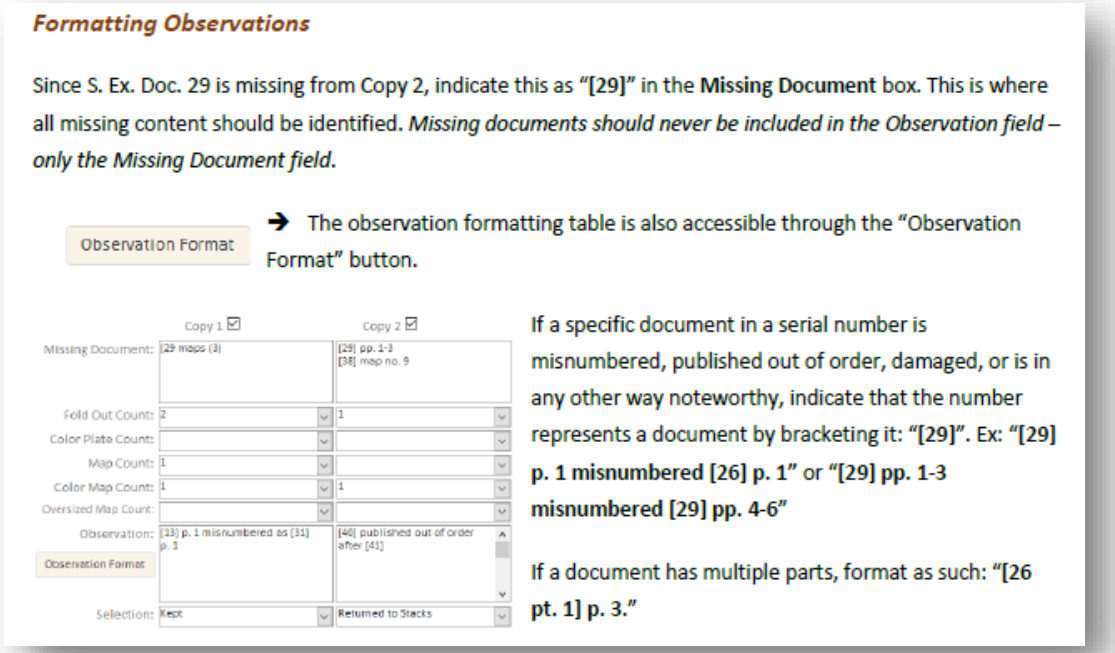
Observation Table

Double-clicking in the observation free text box will retrieve a list of options and their proper format to guide the documentation of inspection notes. *If any content worth noting appears in the volume but not explicitly from the observation choices, note the content verbatim as it appears in the volume.*

Term:	Description:	Example:
[Appendix]	Generally appears at the end of volume. Only include if reflective of whole volume.	**"[Appendix]"**
Barcode	If a volume has a barcode anywhere, list the number following the colon. Will be removed from ILS.	**"Barcode: 1234567890123"**
Book traces	Include handwritten notes or other pencil/pen marks excluding cataloging information.	**"Book traces"**
[Contents]	Or "Table of Contents", or "Table of Documents" – use what the volume lists. Only bracket if reflective of whole volume.	**"[Contents]" "[Table of Contents]"**
Damaged – Acceptable	Indicates imperfect condition of pages and content that do not obscure text and can be digitized. Record type of damage.	**"map no. 1 torn – acceptable"**
Damaged - Unacceptable	Indicates imperfect condition of pages and content that obscures text and illustrations. Record type of damage.	**"map no. 1 in pieces – unacceptable"**

Metadata Guide: Formatting

Formatting metadata observations is especially important to communicate any print or digitization errors among staff and to our vendors.



Formatting Observations

Since S. Ex. Doc. 29 is missing from Copy 2, indicate this as **"[29]"** in the **Missing Document** box. This is where all missing content should be identified. ***Missing documents should never be included in the Observation field – only the Missing Document field.***

Observation Format

➔ The observation formatting table is also accessible through the "Observation Format" button.

	Copy 1 ☑	Copy 2 ☑
Missing Document:	[29 maps (3)	[29] pp. 1-3 [38] map no. 9
Fold Out Count:	2	1
Color Plate Count:		
Map Count:	1	
Color Map Count:	1	1
Oversized Map Count:		
Observation:	[13] p. 1 misnumbered as [31] p. 1	[40] published out of order after [41]
Observation Format		
Selection:	Kept	Returned to Stacks

If a specific document in a serial number is misnumbered, published out of order, damaged, or is in any other way noteworthy, indicate that the number represents a document by bracketing it: **"[29]". Ex: "[29] p. 1 misnumbered [26] p. 1"** or **"[29] pp. 1-3 misnumbered [29] pp. 4-6"**

If a document has multiple parts, format as such: **"[26 pt. 1] p. 3."**

Metadata Guide: Documenting

Metadata is both qualitative and quantitative.

Special Content: Foldouts, Maps, and Plates

Content:	Description:
Foldouts*	Can be charts, plans, and tables that have larger dimensions than standard volume pages.
Two-Page Spread*	Maps or plates that fold out from the center of the volume. Generally found in Signal Officer or Weather Bureau reports.
Maps	Include any geographical content. Usually identifiable by keys, legends, titles, etc. *Color maps have their own count.*
Oversize maps **	Have dimensions larger than 74" x 37". They should count *three* times – 1 foldout, 1 map OR 1 color map, 1 oversized.
Color plates	Plans, charts, tables, illustrations, etc. with any amount of color ink.

Formatting:	Description:
[special content] (#)	Indicates a total # of foldouts, maps, or plates
[#] p. # [special content]	Indicates foldout, map, or plate within a document #/page #
[special content] no. #	Indicates foldout, map, or plate by chronological order – usually in volumes consisting solely of foldouts or singular documents

COUNTING:

* = ***Quantitative data***. Any and all foldout maps (standard and oversized), color maps, and color plates should be counted twice - once as the type of foldout/spread (map, color, or color plate) and once as a foldout OR spread. However, maps, color maps, and color plates are not exclusively foldouts - total the count of standard page maps, color maps, and color plates in addition to that of foldout maps, color maps, and color plates in these fields.

** = Oversize maps count as one oversize, one map, or color map depending on the content. Indicate oversize content when using the database evaluation form.

Adapting to New Challenges

The Serial Set project continues to prioritize evaluation of physical and digital volumes in online and socially distant environments.

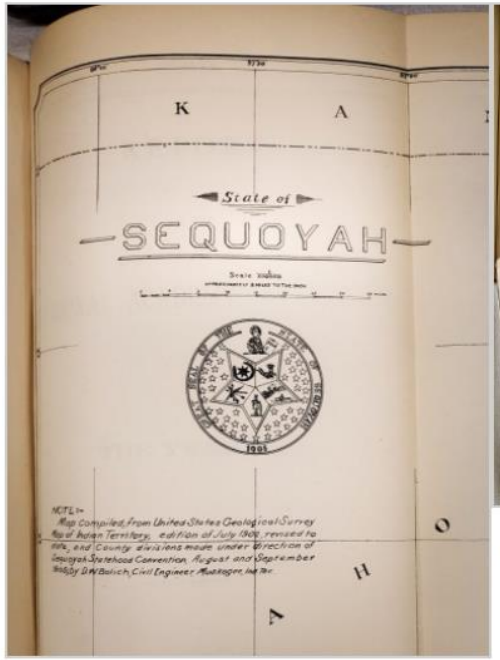
- **Combining onsite and GPO volumes, marking as "RETURN"**
- **Prioritizing Sampler QR (PILOT, and then Full)**
- **Shifting focus to publicity efforts (Story Maps and Blogs)**
- **Coordinating with contractors to pack Ready to Ship volumes and evaluate volumes pulled for blog research**

Outreach – *In Custodia Legis*

- www.blogs.loc.gov/law

While researching further, I discovered a 1906 Senate document entitled "Proposed State of Sequoyah." Senator Joseph B. Foraker of Ohio submitted a draft constitution for Sequoyah, a proposed state for a majority Native American population, to be located in the Oklahoma territory.

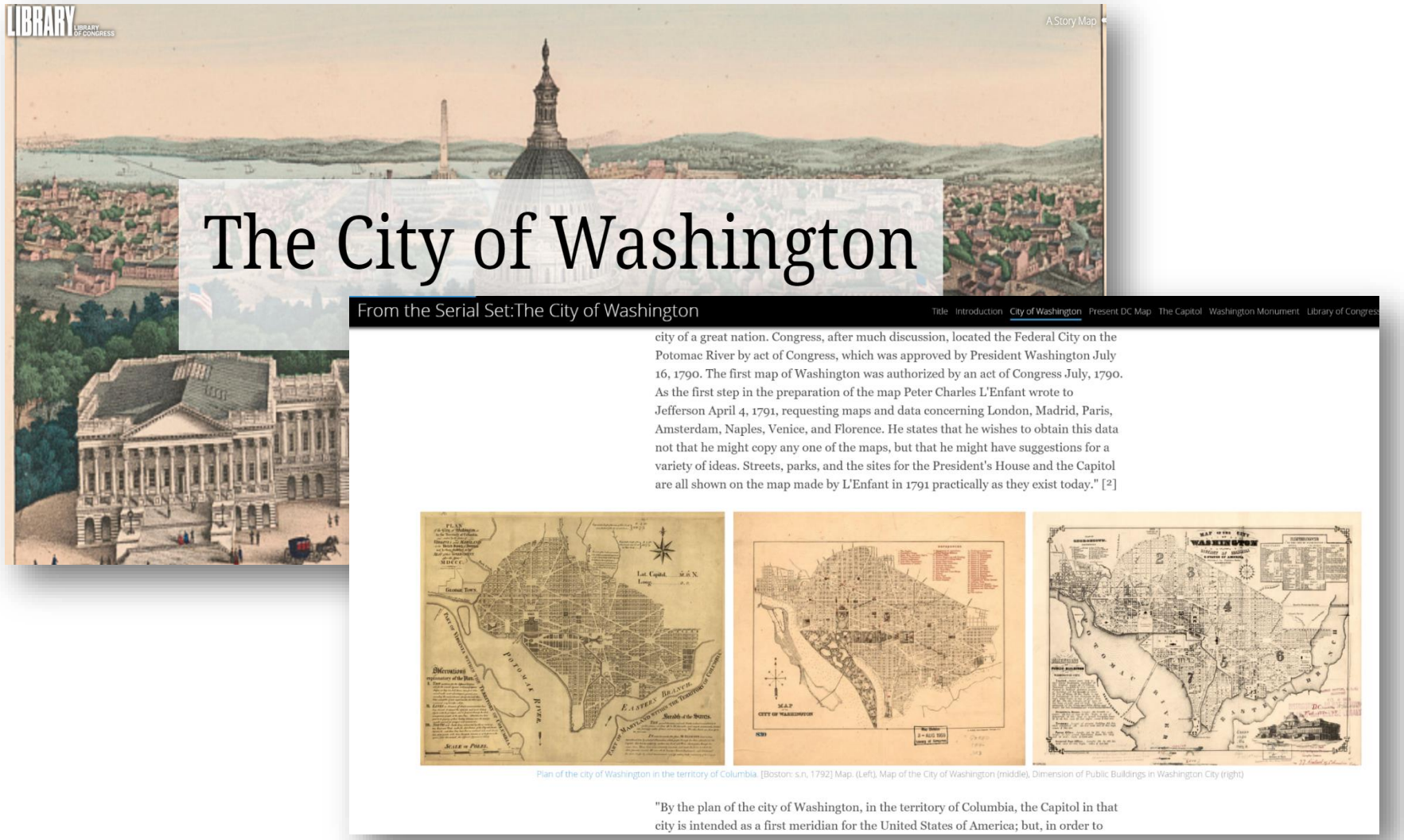
The Serial Set also contains a copy of the map of the proposed state of Sequoyah:



S. Doc. 143, 59th Cong., 1st Sess., unnumbered map (1906) reprinted in Serial Set vol. 4912. Photo by Bailey DeSimone.

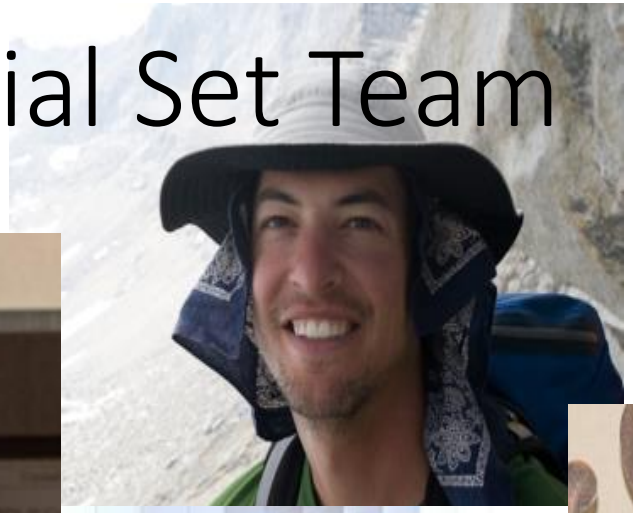
S. Doc. 143, 59th Cong., 1st Sess., unnumbered map (1906) reprinted in Serial Set vol. 4912. Photo by Bailey DeSimone

Upcoming – Story Maps



FEDERAL DEPOSITORY LIBRARY CONFERENCE 2020 #FDLConference

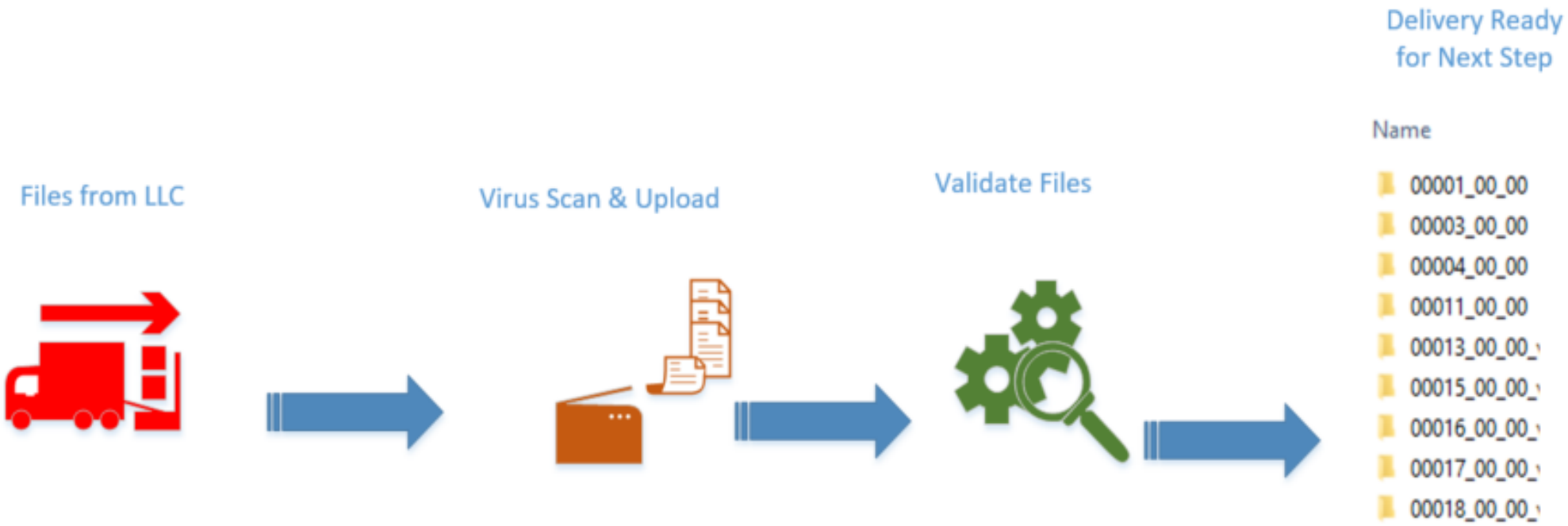
The Serial Set Team



What is GPO doing with the Serial Set?

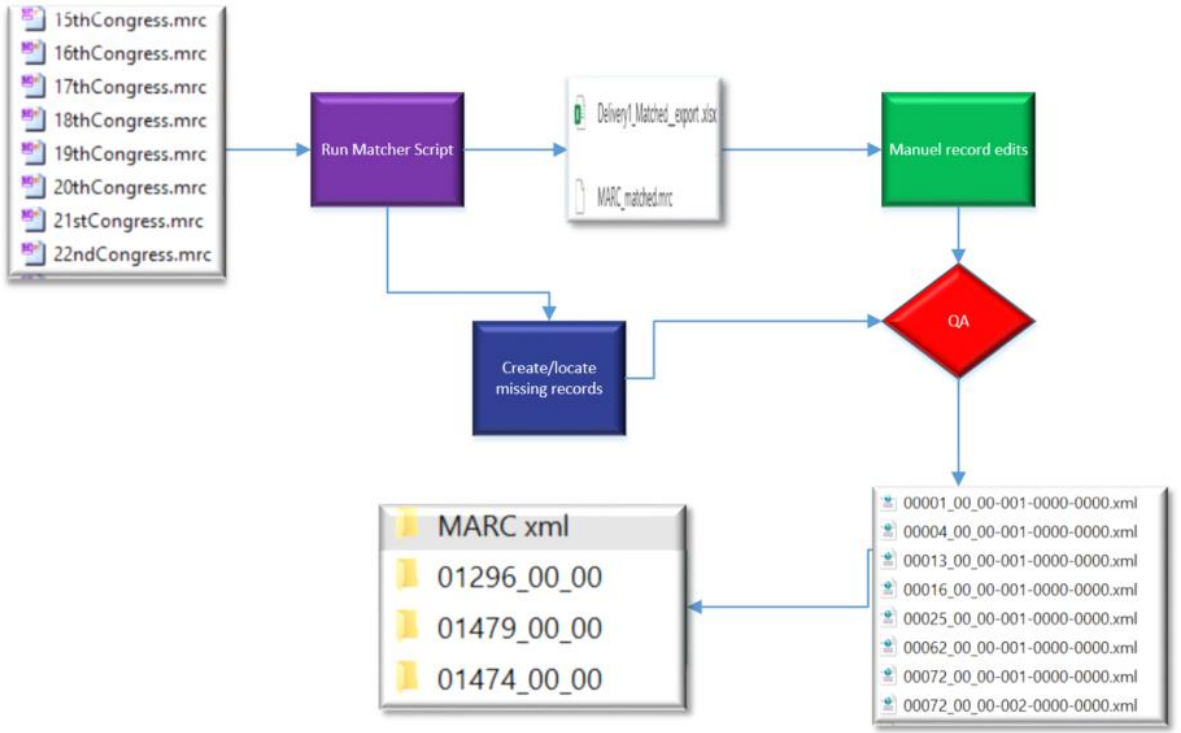
- Acquire Replacement Volumes
- Delivery of Digital files
  - Received this Summer
  - Digital File Receipt Process (Ensuring Integrity)
- Metadata
  - Create/Update Bibliographic Records in the CGP
  - More Metadata!
- Design **govinfo** Serial Set Collection

Receipt of Delivery from LLC – Data Integrity

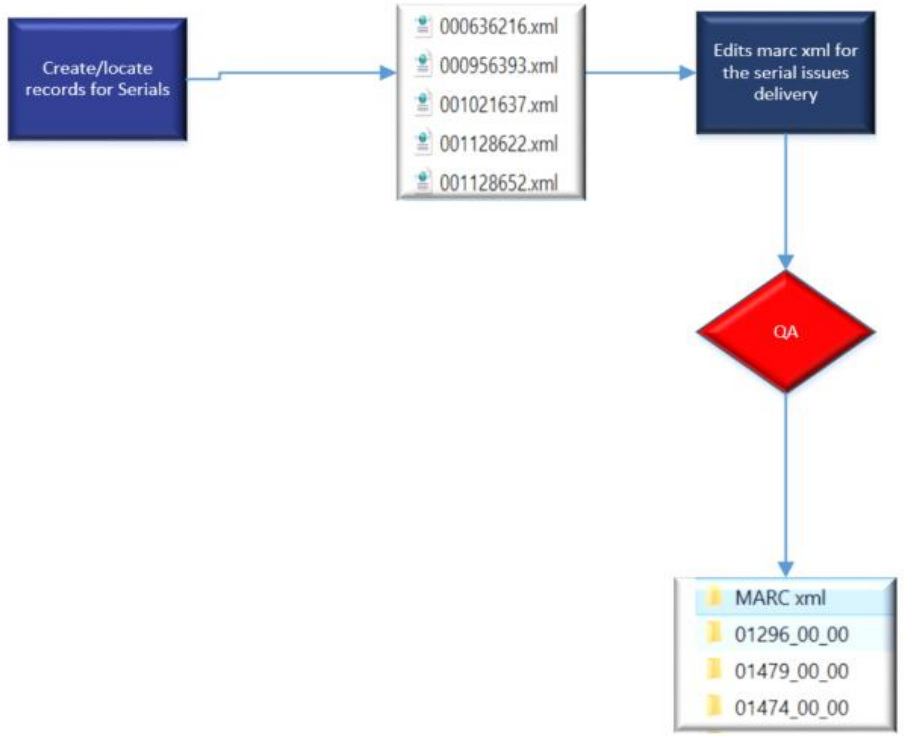


MARC Metadata - Monographs

```
=008  070221s1818\\\\dcu\\\\\os\\\f000\0\eng\d
=040  \\$aWYU$beng$epn$cWYU$dOCLCQ$dOCLCO
=035  \\$a(OCoLC)1065545874
=049  \\$aMAIN
=086  0\$a15-1:S.DOC.181
=086  0\$aY 1.1/2:3
=110  1\$aUnited States.$bCongress.$bSenate.$bCommittee on Pensions.$g(1816-1946)
=245  10$aIn Senate of the United States, April 6, 1818. The Committee on Pensions, to whom w
=260  \\$aWashington, DC,$c1818.
=300  \\$a1 online resource (1 page).
=336  \\$atext$btxt$2rdacontent
=337  \\$acomputer$bc$2rdamedia
=338  \\$aonline resource$bcr$2rdacarrier
=490  1\$aUnited States congressional serial set ;$vserial set no. 3
=490  1\$aSenate document / 15th Congress, 1st session. Senate ;$vno. 181
=500  \\$aTitle from opening lines of text.
=500  \\$aBatch processed record: Metadata reviewed, not verified. Some fields updated by batcl
=590  \\$b15th Congressional serial set
=600  17$aSlocum, Smith P.$2readex congressional thesaurus
=650  07$aClaims.$2readex congressional thesaurus
=650  07$aMilitary pensions.$2readex congressional thesaurus
=650  07$aMilitia and National Guard.$2readex congressional thesaurus
=650  07$aDisabled veterans.$2readex congressional thesaurus
=651  \7$aRhode Island.$2readex congressional thesaurus
=830  \0$aUnited States congressional serial set ;$vserial set no. 3.
=830  \0$aSenate document (United States. Congress. Senate) ;$v15th Congress, 1st session, no. 181.
=922  \\$xSerial_set
=994  \\$a92$bGPO
```



MARC Metadata - Serials

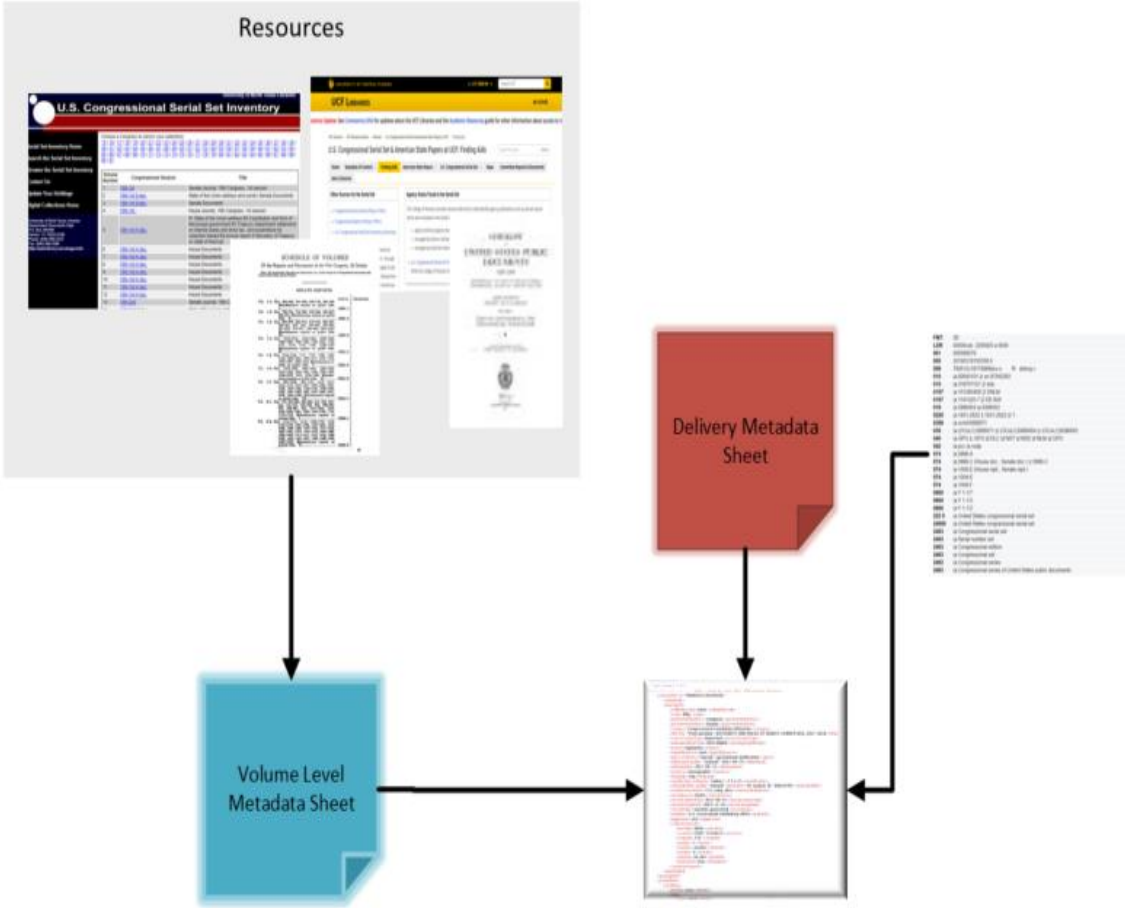



```
    <marc:subfield code="a">1047-A-02 (online)</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="086" ind1="0" ind2=" ">
    <marc:subfield code="a">XJS:15-1</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="110" ind1="1" ind2=" ">
    <marc:subfield code="a">United States.</marc:subfield>
    <marc:subfield code="b">Congress.</marc:subfield>
    <marc:subfield code="b">Senate,</marc:subfield>
    <marc:subfield code="e">author.</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="240" ind1="1" ind2="0">
    <marc:subfield code="a">Journal of the Senate of the United States of America
(Online)</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="245" ind1="1" ind2="0">
    <marc:subfield code="a">Journal of the Senate of the United States of America, First
Session of the Fifteenth Congress.</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="246" ind1="1" ind2=" ">
    <marc:subfield code="a">Senate journal</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="264" ind1=" " ind2="1">
    <marc:subfield code="a">Washington, D.C. :</marc:subfield>
    <marc:subfield code="b">Government Printing Office</marc:subfield>
  <marc:subfield code="c">1818-04-20</marc:subfield>
  </marc:datafield>
```

More Metadata!

- Extract metadata from all available sources
  - Volume Level Master Spreadsheet
  - Delivery Spreadsheet

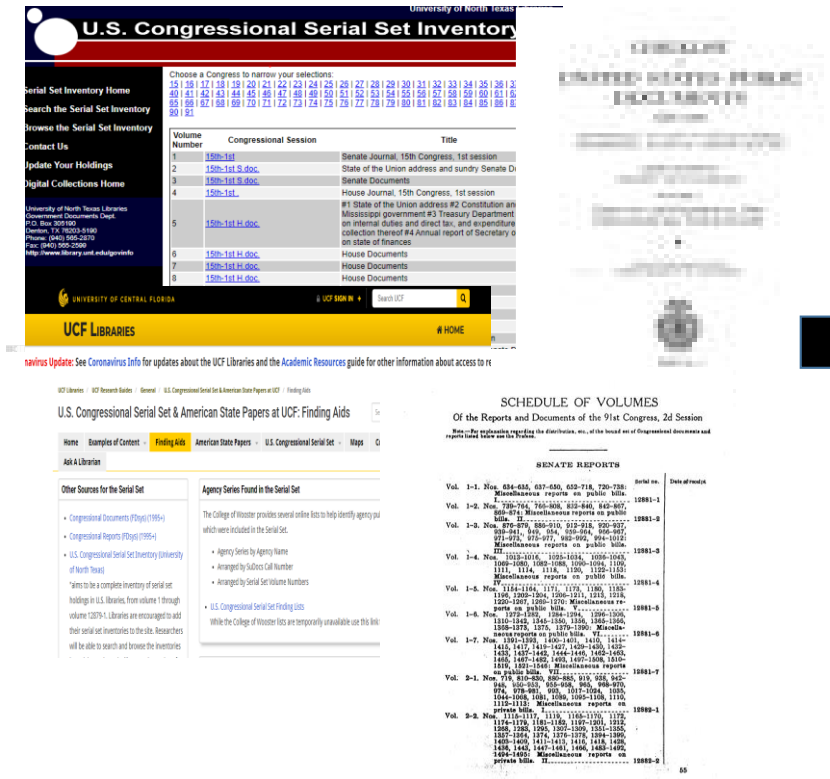
- Generate an initial metadata file that will be submitted to **govinfo** for additional processing



Metadata for govinfo – Bag Delivery Metadata

AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AL	AM	AN	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE
C_PDF_P	LC_PDF_P	LC_PDF_P	LC_PDF_P	LC_PDF_P	LC_PDF_P	LC_PDF_P	D_BatchN	D_Rejecte	Creekside	Note1	Note2	Missing C	QR Status	Feedback	System_L	Serial_Nu	Part	Volume	Congress	Session	Doc_Type	Doc_Title	Index_Par	Index_Vol	Index_Nu	Index_DO	Index_No	Total Num
0001_00_00.pdf:487; 00001_00_00-001-0000-0000.pdf:487							batch_llss_federalistpapers						A			1	0	0	15th	1st	S. journal	Senate Journal			1			
0003_00_00.pdf:571; 00003_00_00-001-0132-0000.pdf:20; 00003_00_0							batch_llss_federalistpapers						A			3	0	0	15th	1st	S. Doc.	Senate Documents	2		3	132-193	Two docs no. 136	
0004_00_00.pdf:607; 00004_00_00-001-0000-0000.pdf:607							batch_llss_federalistpapers						A			4	0	0	15th	1st	H.J.	House Journal			4			
0011_00_00.pdf:630; 00011_00_00-001-0150-0000.pdf:34; 00011_00_0							batch_llss_federalistpapers						A			11	0	0	15th	1st	H.Doc.	House Documents	7		11	149-194	Except 149, not prin	
0013_00_00.pdf:435; 00013_00_00-001-0000-0000.pdf:435							batch_llss_burr						A			13	0	0	15th	2nd	S. journal	Senate Journal			13			
0015_00_00.pdf:329; 00015_00_00-001-0064-0000.pdf:2; 00015_00_00-							batch_llss_burr						A			15	0	0	15th	2nd	S. Doc.	Senate Documents	2		15	64-102	Two docs. No. 68.	
0016_00_00.pdf:443; 00016_00_00-001-0000-0000.pdf:443							batch_llss_burr						A			16	0	0	15th	2nd	H.J.	House Journal			16			
0017_00_00.pdf:750; 00017_00_00-001-0001-0000.pdf:24; 00017_00_0							batch_llss_burr						A			17	0	0	15th	2nd	H.Doc.	State of the Union a	1		17	1-30		
0018_00_00.pdf:690; 00018_00_00-001-0031-0000.pdf:8; 00018_00_00-							batch_llss_burr						A			18	0	0	15th	2nd	H.Doc.	House Documents	2		18	31-48	Except 47, in serial r	
0020_00_00.pdf:552; 00020_00_00-001-0049-0000.pdf:2; 00020_00_00-							batch_llss_burr						A			20	0	0	15th	2nd	H.Doc.	House Documents	4		20	49-85		
0021_00_00.pdf:516; 00021_00_00-001-0092-0000.pdf:516													A			21	0	0	15th	2nd	H.Doc.	#92 On violation of	5		21	92		
0022_00_00.pdf:394; 00022_00_00-001-0086-0000.pdf:12; 00022_00_0							batch_llss_burr						A			22	0	0	15th	2nd	H.Doc.	House Documents	6		22	86-123	Except 92, in serial r	
0023_00_00.pdf:452; 00023_00_00-001-0124-0000.pdf:4; 00023_00_00-							batch_llss_burr						A			23	0	0	15th	2nd	H.Doc.	House Documents	7		23	124-136		
0024_00_00.pdf:186; 00024_00_00-001-0137-0000.pdf:27; 00024_00_0							batch_llss_herculesmulligan						A			24	0	0	15th	2nd	H.Doc.	House Documents	8		24	137-150		
0025_00_00.pdf:499; 00025_00_00-001-0000-0000.pdf:499							batch_llss_herculesmulligan						A			25	0	0	16th	1st	S. journal	Senate Journal			25			
0028_00_00.pdf:112; 00028_00_00-001-0099-0001.pdf:112							batch_llss_herculesmulligan						A			28	0	0	16th	1st	S. Doc.	#99 Abstracts of all	3		28	99	Unpaid bonds for d	

Metadata for govinfo – Volume Level Metadata



Titles

Document Numbers

Serial number

Congress

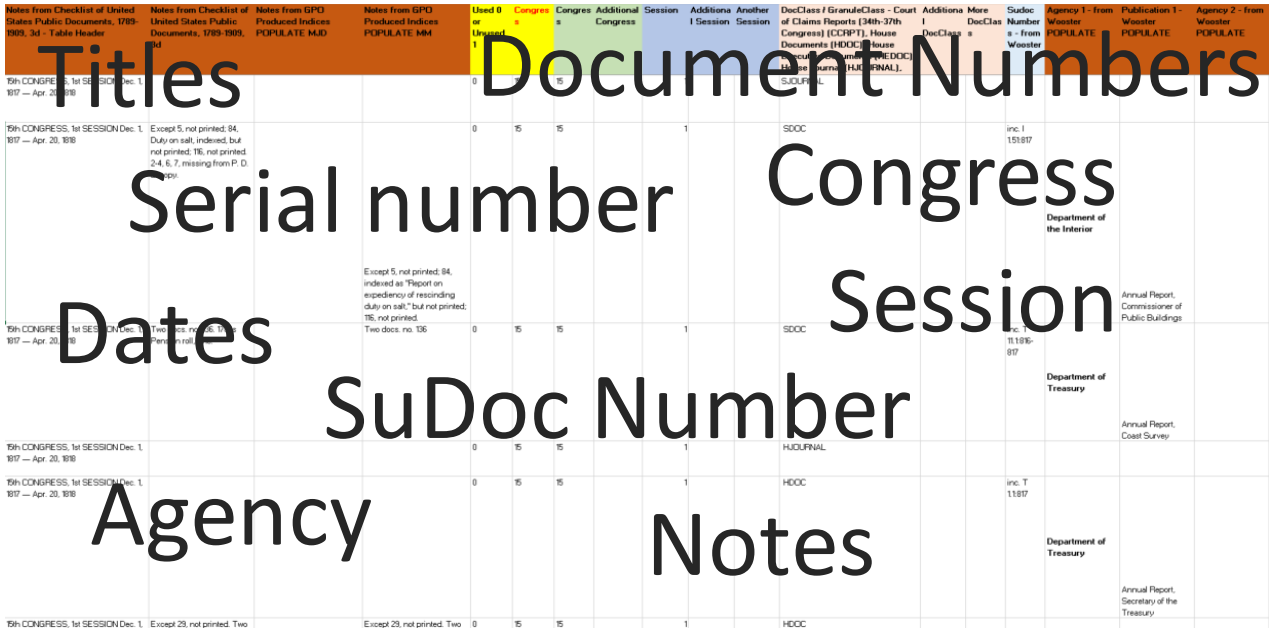
Dates

Session

SuDoc Number

Agency

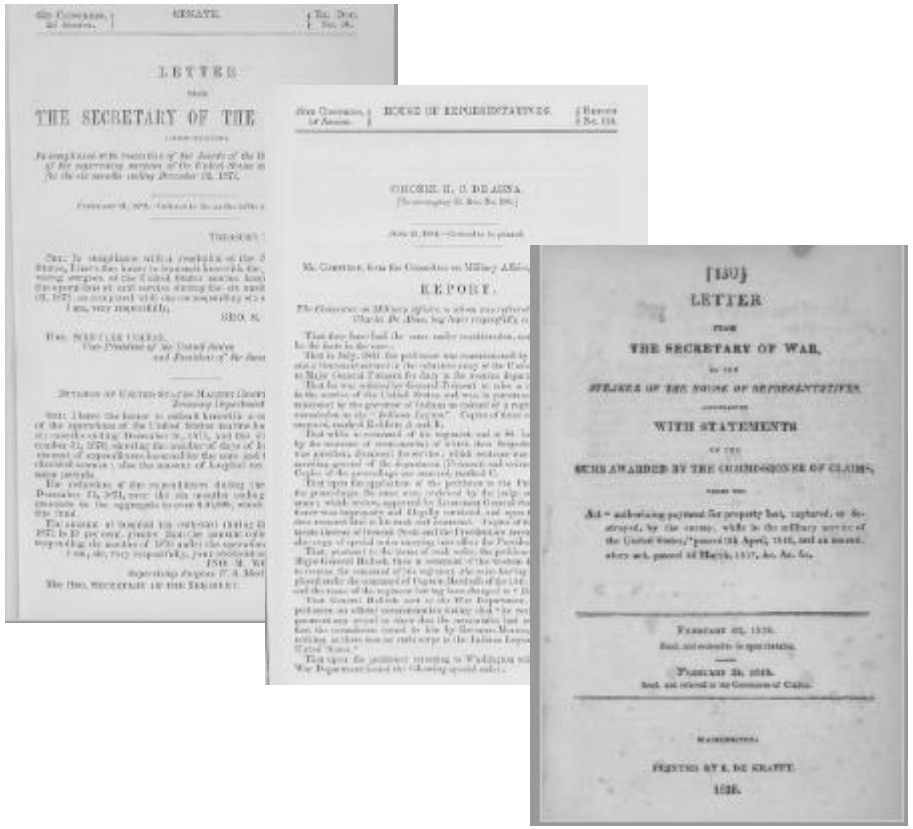
Notes



What's Next?

- Workflow Development and Process Validation
- Continue Metadata Creation, Review, and Gap Analysis
- Process Analysis and Collection Discovery for **govinfo**
- Serial Set Collection Design for **govinfo**

govinfo Serial Set Collection Release Date



- Fall 2021: First release of content
- Additional volumes to follow as processed
- Completion of collection will take several years

Keep Up-to-Date on the Project

- Updates on FDLP.gov
- Needed Serial Set Volumes
- From the Serial Set articles on the Law Library blog

Questions?