NOTES ON OPERATIONS

Ebook Call Number Project

Automating Library of Congress Classification Numbers to Ebooks for Reporting

Bonita Pollock, Zhao Liu, Amanda Yesilbas, Laura Pascual, and Skyler Schultz

The Ebook Call Number Project at the University of South Florida Libraries automated the process of adding Library of Congress (LC) Classification numbers into call number fields, thereby enhancing ebook metadata while minimizing manual work. This initiative aimed to improve the accuracy of analytics reporting for discipline-specific collection assessment and reporting. The project involved analyzing existing metadata, developing a Python-based automation tool using OCLC's Connexion API, and implementing a systematic workflow within the Ex Libris Alma Library Management System for cataloging and quality control. The effort increased the number of ebook records with LC call numbers from 91 percent to 99 percent, significantly enhancing reporting accuracy and efficiency. This article details the project's rationale, methodology, and outcomes, demonstrating how automation and artificial intelligence can streamline metadata enrichment in academic libraries. Finally, the article presents an alternative method for smaller institutions that lack access to programmers or programming solutions and discusses future applications.

Introduction

The University of South Florida (USF) Libraries completed an extensive project that enhanced ebook metadata by inserting Library of Congress (LC) Classification (LCC) into the call number field, allowing staff to produce more accurate and detailed analytical reports on the USF Libraries' collections by subject area. The Metadata/Cataloging Unit of the Collections & Discovery (C&D) Department needed to analyze the electronic resource records and determine how many vendor records contained LC call numbers. The unit devised a plan to incorporate LC call numbers into as many electronic records as possible to improve the results of Alma Analytics reports. This article discusses why USF Libraries chose to undertake this enhancement project and details the steps involved, including planning, implementation, and outcomes.

Bonita Pollock (pollockb1@usf.edu) is Associate Director, Collections & Discovery, University of South Florida Libraries, https://orcid.org/0000-0001-5500-6862. **Zhao Liu** (zhaoliu@usf.edu) is Electronic Resources Catalog Librarian, University of South Florida Libraries, https://orcid.org/0009-0000-0685-7193. **Amanda Yesilbas** (ayesilbas@usf.edu) is E-Resources Librarian, University of South Florida Libraries, https://orcid.org/0009-0007-2348-2312. **Laura Pascual** (lcpascua@usf.edu) is Director Collections & Discovery, University of South Florida Libraries, https://orcid.org/0000-0002-9565-4466. **Skyler Schultz** (skyler21@usf.edu) is Complex Copy Cataloger, University of South Florida Libraries, https://orcid.org/0009-0009-0075-0227.

Background

The USF is a large public R1 research institution with nearly 50,000 students. Under a single accreditation, the USF has locations in Tampa, St. Petersburg, and Sarasota. USF is a metropolitan institution, integrated with the local community, and diverse. In 2023, the USF joined the prestigious Association of American Universities (AAU). The USF Libraries are composed of the Tampa Campus Library, the Nelson Poynter Library in St. Petersburg, and the Library Services and Information Commons on the Sarasota campus. The University Health Libraries consists of two libraries: the Shimberg Health Sciences Library on the Tampa campus and the Florida Blue Health Knowledge Exchange in downtown Tampa.

In 2021, the twelve universities in the State University System (SUS) and the twenty-eight state colleges migrated to the Ex Libris Alma library management system and the Primo VE discovery tool. On a consortial level, the library system and discovery tool are managed by the consortium; however, the USF Libraries C&D Department manages the bibliographic content and the discovery interface for all USF Libraries. Cataloging practices are coordinated across all campuses.

The Tampa Campus Library has adopted an electronic-first acquisition strategy since 2009. It directly manages five evidence-based ebook acquisition programs and one very large ebook data-driven acquisitions program. Most electronic content is available across all USF campuses and colleges through a single shared authentication process with common licensing. The Tampa Library is working to make better use of Alma functionality, using it as a repository for all collections-related data to provide a cohesive source for reporting. If Alma has good data, good reporting can be developed from it. This includes using Alma Acquisitions for purchase information and SUSHI harvesting to Alma for COUNTER R5 usage report data.

It is essential for all academic libraries to report and analyze their content, whether for developing and managing the collection, demonstrating the value of the library's collections, providing subject-area support information, preparing accreditation reports, developing new program proposals, securing grant support, or benchmarking, among other purposes. Developing an automated, standardized reporting process requires a consistent, reliable data source with key metrics. This case study specifically addresses the need to report USF Libraries' electronic collections metrics on inventory and usage by subject area, using the LCC in the MARC bibliographic call number fields.

Aside from reporting for collection management, USF Libraries need to report externally for a variety of purposes, including those that relate to demonstrating the value of the library's collections, for letting colleges and departments know what content the library has that supports a given discipline, for college program accreditations, for grant support, and for benchmarking. Reporting required at the USF Libraries included the Integrated Postsecondary Education Data System (IPEDS) and the Association of College and Research Libraries (ACRL), as well as the library support sections of program accreditation applications and new degree programs. Because the basis for most of this reporting is the subject area or discipline, the LCC code data fields—using standardized alphanumeric coding—are an excellent way

to report content by subject. The USF Libraries prepare annual reports that include specific assessments of collections and provide data to meet the USF's accreditation requirements. One such document, the "Report Card," was devised to report library collection content and usage data by discipline.

Literature Review

The LCC system functions as a foundational organizational and analytical structure for academic library collections, shaping collection development practices for decades. Historically, LCC enabled libraries to describe, measure, and evaluate their print holdings through standardized subject ranges and subclass alignments, thereby providing the basis for quantitative and qualitative assessment. Early methodologies such as the RLG Conspectus and the North American Collections Inventory Project relied on LCC's disciplinary structure to determine collecting strengths, define peer comparisons, and assess institutional priorities.[1] Contemporary iterations, including OCLC's Conspectus-based WorldShare Collection Evaluation tools, continue to employ LCC ranges to support benchmarking and historical analysis.[2]

LCC has also been central to quantitative measurement approaches that use call number boundaries to define countable segments of the collection. Shelf-list counts and comparative subject inventories are widely used to determine relative subject strengths, collecting intensity, and collection gaps.[3] Targeted evaluation practices, including the Brief Test methodology, similarly rely on LCC as a validating mechanism to define the scope of a subject study and exclude materials that fall outside the designated subclass range.[4] Together, these approaches demonstrate LCC's long-standing role as a shared analytical language for academic library assessment.

Beyond counting and benchmarking, LCC supports the analysis of usage and curricular content. Circulation statistics aggregated by call number ranges allow researchers to evaluate subject demand, identify fields that are overused or underused, and inform purchasing decisions.[5] This granular usage mapping clarifies how collections interact with institutional teaching and research needs, particularly when paired with curricular mapping based on LCC subclasses.[6] These methods provide evidence that call number segmentation produces focused, discipline-specific insights that broad subject labels alone cannot supply.

Researchers acknowledge, however, that the structure of LCC also introduces persistent assessment barriers. Because LCC organizes knowledge primarily by discipline, interdisciplinary subjects can become fragmented across multiple classification ranges, thereby complicating systematic evaluation. Research on Women's Studies collections, for example, illustrates how LCC distributes content across subclasses, hindering accurate analysis and complicating comparative evaluation.[7] This core structural constraint continues to limit the precision of LCC-based assessment even in the digital era.

The growth of electronic monograph collections has expanded LCC's operational scope. Although ebooks are not physically shelved, librarians increasingly rely on call numbers as metadata to manage digital content, map subject overlap, and support the development of user interfaces.[8] Tools such as the Hierarchical Interface to Library of Congress Classification (HILCC) demonstrate how call number data can support browsable classification systems for digital collections.[9] Large-scale ebook mapping

projects, such as those at the University of North Texas, have implemented LCC ranges to reflect overlapping and interdisciplinary relationships across digital packages.[10] These developments suggest that LCC retains value not only as a structural classification scheme for print formats but also as a metadata framework for digital resource discovery and assessment.

At the same time, LCC-based ebook assessment introduces notable challenges. Many electronic records lack call numbers because cataloging standards historically did not require classification for nonphysical formats.[11] Retrospective call number assignment and metadata enhancement require substantial labor investment, and topic-mapping techniques that connect LCC subclasses to subject categories demand continuous refinement.[12] Moreover, the disciplinary bias that fragments interdisciplinary subjects in print collections remains equally problematic in the digital environment.[13] Nevertheless, ebook usage studies show the growing value of LCC for aggregating platform-specific data—such as section requests and page views—into meaningful subject-level analytics.[14]

Several studies have examined the automation of the creation of LCC in MARC records. In the White Paper "Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper," authors Suzanne Kemperman et al. indicated that automation in bibliographic data management plays an important role in enhancing efficiency and accuracy.[15] According to a study conducted by Isidoro Gil-Leiva et al. on South American libraries, almost 59 percent of the respondents surveyed felt that using automation to assign classification codes was appropriate.[16] This research provided the team with strong motivation to develop an automated process for creating LC call numbers in catalog records.

Taken together, the literature indicates that LCC continues to function as a relevant and adaptable framework for academic collection assessment. Although its disciplinary structure and metadata gaps pose methodological limitations, LCC remains a critical tool for subject-based analysis, curricular alignment, usage interpretation, and digital collection mapping. As electronic monographs grow in scale and complexity, LCC's metadata utility positions the system not as a relic of print organization but as an evolving data structure capable of supporting evidence-based collection development in hybrid print-digital environments.

Report Cards

In the spring of 2023, the C&D Department at the USF Libraries was asked by library administration to create collection report cards for the USF Libraries' annual report. These report cards allowed USF Libraries to communicate which library resources are available in specific disciplines to departments across the campuses. Library administration wanted the report cards to be a one-page snapshot of subject-level collections and usage data, aligned with academic disciplines and colleges/departments. Each report card focused on a specific discipline at the USF and highlighted the libraries' resources in print and electronic formats for that subject area, according to the LCC. Therefore, inventory by LC call number and usage/circulation by LCC for print and electronic books and journals needed to be in Alma. Initially, no working reports were found, and it did not appear that LC call numbers were present on a substantial portion of the records in critical discipline areas.

The decision to use the LCC to categorize library resources was twofold. First, LCC is more streamlined and easier to manage than LC Subject Headings. LC Subject Headings exhibit greater variation and combinations, making subject heading classification and searching more difficult. In addition, in her book *Managing E-Book Metadata in Academic Libraries: Taming the Tiger*, Donna E. Frederick stated that classification should be included in the ebook metadata process to assist patrons with selection by topic when using electronic shelf listing.[17] Second, Alma Analytics already includes LCC in its reporting, making it more convenient to use when creating custom reports. The Ebook Call Number Project Team decided to focus on LCC for ebooks first and then branch out to e-journals for a future project.

In the first year, report cards were created manually. This manual process was developed using vendor title lists or entitlements reports, vendor MARC records, Alma electronic collection exports, COUNTER reports, Alma circulation data, and physical item counts. It was difficult to determine the LCCs from the vendor inventory title lists because many appeared to be missing. It was also difficult to then manually match those titles to the COUNTER reports to get COUNTER usage by LCC. Exported Alma electronic collections for book and journal platforms also appeared to be missing many LC call numbers. With this manual process for ebooks and e-journals, the data was limited to major vendors with COUNTER reports pertinent to the subject area. This informed the goal of improving the data in Alma to enable more automated report card production. Achieving this goal and facilitating future report card creation required enhancing all electronic records in Alma with SUSHI COUNTER statistics and LCC numbers.

The top section of figure 1 contains Program Area data that is broader and created from keywords in subject searches using publication date as a limiter. This yields a year-over-year comparison of inventory only. The middle and bottom left sections of the Snapshot of Subject Area Holdings portion of figure 1 are created from inventories by LCC, however, and include usage data. Figure 2 is a close-up of the lower left section of the report card highlighting the Print Inventory by Subject Area and the Electronic Inventory by Subject Area with numbers of titles with usage and usage for books and journals for the selected LC subclasses. This project concerns the Electronic Inventory by Subject Area section. The pie charts in figure 1 were derived from this data.

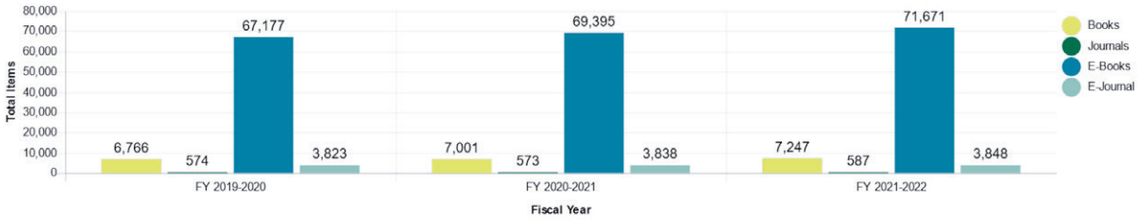
Project Overview

In the summer of 2023, the Metadata/Cataloging Unit created an Ebook Call Number Project Team and tasked the group with improving the process of creating reports via enhancing the electronic records. Although many members of the C&D Department provided input, the core team consisted of nine members: the head of metadata/cataloging, an e-resource catalog librarian, an electronic resources librarian, an applications developer programmer, a metadata specialist, a metadata quality-control manager, and three complex catalogers. The bulk of the work was completed by four key members who organized the project, ran the workflows, and recorded results. This extensive project started in the summer of 2023 and concluded in the spring of 2024. The team spent three months planning, longer than six months in implementation, and two months building reports.

ENGINEERING

TRENDS OVER 3 FISCAL YEARS (FY 2019-2022)

Trend of Format Holdings by Program Area Per FY



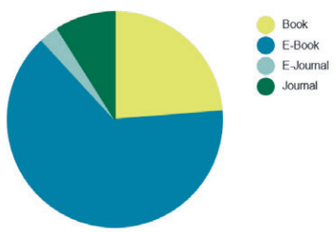
Program Area	Books	E-Books	Journals	E-Journal
FY 2019-2020	**6,766**	**67,177**	**574**	**3,823**
Engineering	5,446	46,768	393	2,703
Civil Engineering	193	1,575	29	180
Mechanical Engineering	230	3,757	42	223
Electrical Engineering	142	6,642	50	289
Materials Engineering	369	4,034	26	213
Chemical Engineering	260	2,059	33	180
Management Systems Engineering	126	2,342	1	35

Program Area	Books	E-Books	Journals	E-Journal
FY 2020-2021	**7,001**	**69,395**	**573**	**3,838**
Engineering	5,575	48,300	393	2,738
Civil Engineering	188	1,669	29	181
Mechanical Engineering	260	3,914	42	224
Electrical Engineering	142	6,747	42	262
Materials Engineering	384	4,227	26	218
Chemical Engineering	281	2,139	33	184
Management Systems Engineering	171	2,399	8	31

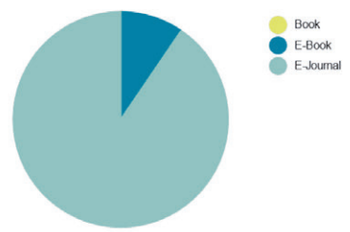
Program Area	Books	E-Books	Journals	E-Journal
FY 2021-2022	**7,247**	**71,671**	**587**	**3,848**
Engineering	5,656	49,795	393	2,746
Civil Engineering	229	1,747	29	181
Mechanical Engineering	274	4,088	42	225
Electrical Engineering	163	6,847	42	262
Materials Engineering	408	4,461	40	218
Chemical Engineering	383	2,251	33	185
Management Systems Engineering	134	2,482	8	31

SNAPSHOT OF SUBJECT AREA HOLDINGS (FY 2021-2022)

Collection Holdings by Format



Collection Usage by Format **



High Impact Journals

Journal Title	Impact Factor	USF Holding
Energy & Environmental Science	39.714	☑
Progress in Energy and Combustion Science	35.339	☑
Nature Electronics	33.255	☐
Nature Biomedical Engineering	29.234	☐
Applied Catalysis B-Environmental	24.319	☑
IEEE Transactions on Pattern Analysis and Machine Intelligence	24.314	☑
Advances in Applied Mechanics	17.400	☑
Bioactive Materials	16.874	☑
Chemical Engineering Journal	16.744	☑
Biomaterials Research	15.863	☑
BIOMATERIALS	15.304	☑
IEEE Signal Processing Magazine	15.204	☑
Proceedings of the IEEE	14.910	☑
npj Regenerative Medicine	14.404	☑
IEEE Transactions on Neural Networks and Learning Systems	14.255	☑
Journal of Hazardous Materials	14.224	☑
Engineering	12.834	☑
Journal of Industrial Information Integration	11.718	☑
Additive Manufacturing	11.632	☑
Automation in Construction	10.517	☑

Print Inventory by Subject Area

Subject	Print: Inventory	Print: Uses	Print: Percent Use	Print: Titles Used	Journal: Inventory
General & Civil Engineering	3,786	65	22%	55	2,255
Environmental Engineering	2,461	42	14%	25	1,058
Mechanical Engineering	1,381	40	14%	26	384
Electrical/Computer Engineering	5,523	97	33%	64	1,123
Chemical & Materials Engineering	1,592	35	12%	34	1,047
Industrial/Manufacturing Systems Engineering	1,959	14	5%	10	432
TOTAL:	**16,702**	**293**	**100%**	**214**	**6,299**

Electronic Inventory by Subject Area

Subject	E-Book: Inventory	E-Book: Uses	E-Book: Percent Use	E-Book: Titles Used	E-Journal: Inventory	E-Journal: Uses	E-Journal: Percent Use	E-Journal: Titles Used
General & Civil Engineering	13,854	4,115	26%	1,346	404	29,759	19%	334
Environmental Engineering	3,475	1,932	12%	430	71	18,268	12%	63
Mechanical Engineering	4,602	1,980	12%	364	117	9,782	6%	96
Electrical/Computer Engineering	15,572	4,945	31%	1,340	186	11,515	7%	157
Chemical & Materials Engineering	4,298	2,201	14%	576	249	33,403	22%	208
Industrial/Manufacturing Systems Engineerin	2,277	403	3%	197	61	4,526	3%	49
Engineering - General	1,174	427	3%	88	948	47,592	31%	431
TOTAL:	**45,252**	**16,003**	**100%**	**4,341**	**2,036**	**154,845**	**100%**	**1,338**

Subject Area Databases

Database Title
Scopus
Web of Science
Science Direct
Springerlink
IEEE Xplore
ACM Digital Library
Applied Science & Technology Source
Inspec

For Collection Usage by Format: Usage counts are not available for print journals. For Inventories: Subject areas differ by format. High Impact Journal data was collected from Journal Citation Reports. ** For the "Collection Usage by Format" pie chart: the small numerical value for the "Book" format wedge makes it not visible. Please consult the inventories for value comparisons.

Figure 1. Example of a report card for engineering.

Print Inventory by Subject Area

Subject	Print: Inventory	Print: Uses	Print: Percent Use	Print: Titles Used	Journal: Inventory
General & Civil Engineering	3,786	65	22%	55	2,255
Environmental Engineering	2,461	42	14%	25	1,058
Mechanical Engineering	1,381	40	14%	26	384
Electrical/Computer Engineering	5,523	97	33%	64	1,123
Chemical & Materials Engineering	1,592	35	12%	34	1,047
Industrial/Manufacturing Systems Engineering	1,959	14	5%	10	432
TOTAL:	**16,702**	**293**	**100%**	**214**	**6,299**

Electronic Inventory by Subject Area

Subject	E-Book: Inventory	E-Book: Uses	E-Book: Percent Use	E-Book: Titles Used	E-Journal: Inventory	E-Journal: Uses	E-Journal: Percent Use	E-Journal: Titles Used
General & Civil Engineering	13,854	4,115	26%	1,346	404	29,759	19%	334
Environmental Engineering	3,475	1,932	12%	430	71	18,268	12%	63
Mechanical Engineering	4,602	1,980	12%	364	117	9,782	6%	96
Electrical/Computer Engineering	15,572	4,945	31%	1,340	186	11,515	7%	157
Chemical & Materials Engineering	4,298	2,201	14%	576	249	33,403	22%	208
Industrial/Manufacturing Systems Engineering	2,277	403	3%	197	61	4,526	3%	49
Engineering - General	1,174	427	3%	88	948	47,592	31%	431
TOTAL:	**45,252**	**16,003**	**100%**	**4,341**	**2,036**	**154,845**	**100%**	**1,338**

Figure 2. Report card data sheet.

Action Plan

The team developed an action plan for the Ebook Call Number Project that included collecting call number data for electronic resources in Alma, experimenting with methods to add call numbers to records in Alma in batches, exploring ways to retrieve and insert call numbers into records, developing an automation workflow for completing the call number retrieval and insertion, collecting statistics on the project, and creating custom analytics reports to make report cards.

Phase one of the action plan was to collect data from Alma, the Library Services Platform. A team member conducted an extensive survey of all the electronic ebook collections activated in Alma for the University. The collections were exported as MARC records and analyzed in MarcEdit. Field counts were tracked for LC call number, author, title, and subject across vendors and ebook collections in an Excel spreadsheet. Figure 3 shows a sample tracking spreadsheet.

The analysis revealed that in the spring of 2023, USF Libraries had seventy-five ebook collections activated in Alma's Community Zone (CZ) from twenty-eight different vendors. Filters were used to exclude primary source materials, government documents, and other material types, leaving only books. Approximately eighteen of the electronic resource vendors already provide LCC call numbers on 90 percent or more of their records. These collections consisted mainly of resources from major ebook vendors. Figure 4 shows vendor counts highlighting the percentage of call numbers present in the MARC records.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Vendor	Collection	Alma Electronic Collection Name	Portfolio Count	MARC Total Count(from MarcEdit)	Call Number Percentage 050/total	035	050	090	100	110	111	130	245	264	505	520	650	830	Notes
26	ACLS	ACLS Humanities ebook	ACLS Humanities eBook	5962	5924	87.68%	5924	5194	2128	5264	33	9	37	5924	2565	2961	2389	5357	3372	
27	ABC-Clio	ABC-Clio ebooks	ABC-CLIO ebooks	339	339	87.91%	339	298	0	269	0	0	0	339	78	243	186	303	110	
28	Cambridge	Cambridge EBA	Cambridge EBA (USF)	47283	47283	89.83%	47283	42474	338	35535	96	333	93	47283	47257	23146	47126	39125	22141	5/5/2(
29	EBSCO	owned	EBSCOhost Ebooks	68708	68701	89.96%	68701	61803	1016	51308	2232	426	140	68701	29245	39114	35773	59750	16312	no marc record for collectio
30	Project Muse	owned	Project Muse All Books	6160	6149	92.03%	6149	5659	54	5090	20	25	24	6149	4609	3970	4401	5773	2771	no marc record for collectio
31	LOEB		Loeb Classical Library	641	292	92.12%	292	269	0	256	0	0	5	292	80	152	154	202	272	no marc record for collectio
32	JSTOR	ebooks owned	JSTOR Books	6514	6365	92.38%	6365	5880	7	5427	34	49	24	6365	5274	3107	4883	6020	2493	no marc record for collectio
33	Brill	Brill EBA	Brill Ebooks EBA (Brill - All E-Book Titles)	27172	27155	93.32%	27155	25342	115	16731	442	323	145	27155	20917	21537	22290	23693	21433	5/5/2(
34	Wiley	Online Library ebooks	Wiley Online Library - AutoHoldings Books	6349	6349	93.38%	6349	5929	209	3292	88	74	55	6349	3836	4538	3732	5879	2157	
35	Knovel	Knovel Civil Engineering & Construction Materials Academic	Knovel Civil Engineering & Construction Materials Academic	1108	1108	93.68%	1108	1038	6	561	302	96	4	1108	926	411	638	1065	290	
36	Talor & Francis	Owned/EBA Autoload	Taylor & Francis eBooks Complete	36288	35610	94.24%	35610	33559	208	23921	108	140	107	35610	32661	19337	22710	33342	15532	5/5/2(
37	Brill	Brill EBA	EBA Brill All	22890	22878	94.68%	22878	21661	49	13968	425	264	145	22878	18077	18352	18295	19785	19386	5/5/2(
38	Ebook Central	Owned	Ebook Central Perpetual, DDA and Subscriptiion Titles	246012	243581	94.81%	243581	230930	5945	169656	11979	5082	630	243581	160681	174940	161060	222245	100974	5/5/2(
39	Knovel	Knovel Electrical & Power Engineering Academic	Knovel Electrical & Power Engineering Academic	487	486	94.86%	486	461	15	314	62	7	3	487	336	249	260	469	164	
40	Knovel	Knovel Mechanics & Mechanical Engineering Academic	Knovel Mechanics & Mechanical Engineering Academic	1062	1062	95.10%	1062	1010	16	693	190	20	4	1062	803	413	584	1040	246	
41	Sage		Sage (USF)	110	110	95.45%	110	105	5	9	0	0	8	110	62	2	99	101	1	no marc record for collectio
42	Elsevier (Science Direct platform)	Elsevier ScienceDirect Books (Elsevier ScienceDirect Books Complete)	Elsevier ScienceDirect Book Series Pa ckage - All Subjects	5689	5689	96.59%	5689	5495	85	1308	27	30	52	5689	3124	3776	3370	5587	3078	
43	IEEE Xplore platform	IEEE Xplore Wiley	IEEE Xplore Wiley eBooks Library (IEEE Xplore)	1175	1159	96.81%	1159	1122	8	878	0	2	1	1159	1145	1069	985	1125	559	

Figure 3. Data collection spreadsheet: analysis of ebook call numbers.

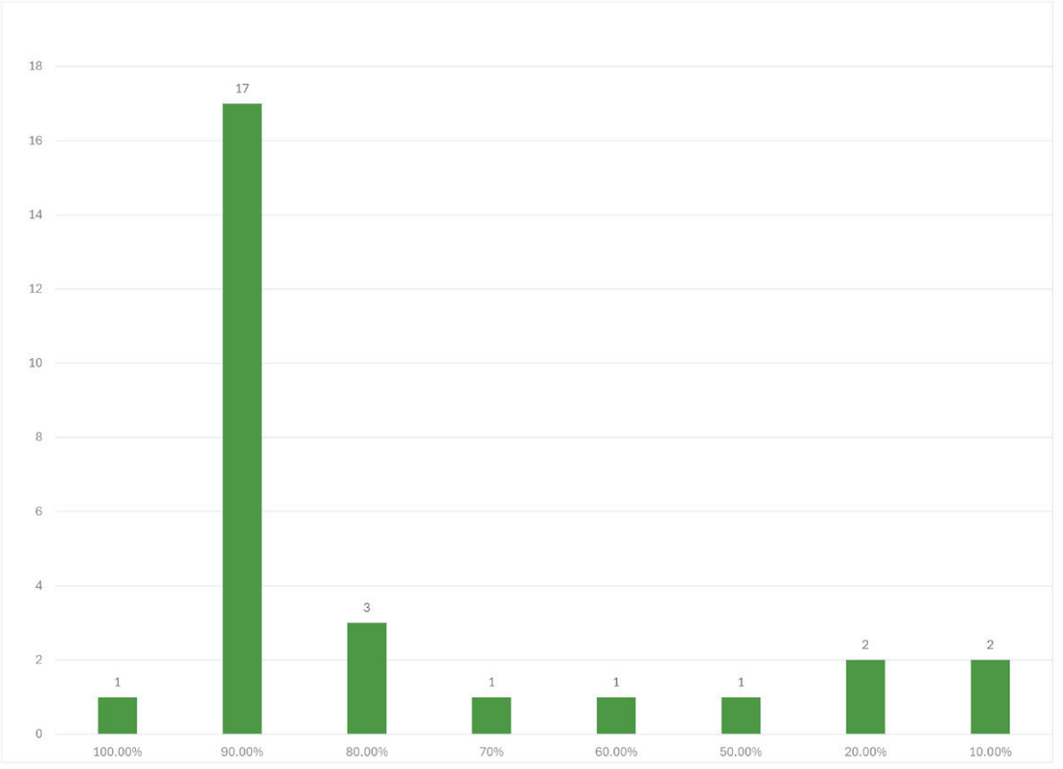


Figure 4. Vendor counts by call number percentage.

The remaining ten vendors were mostly smaller or specialized publishers. These vendor collections required more call number enhancements. A total of more than 60,000 ebook records without LC call numbers required remediation in the catalog. These records represented critical academic disciplines,

including foreign language materials, electrical engineering, psychology, archeology, medieval history, reference materials, and music and drama collections. It was important for USF Libraries to represent these disciplines accurately in the report cards. Assigning call numbers to these collections added clarity and completeness while improving the accuracy of analytics call number reporting, including those used on the report cards. Thus, staff time and effort invested in this project yielded valuable results.

The Ebook Call Number Team proceeded to the second phase of the action plan after determining that it would be useful for USF Libraries to add LCC to all ebooks. In this phase, the team experimented with various methods to batch add call numbers to MARC records in Alma. Most of the USF Libraries' electronic resources were activated in Alma's CZ, creating an Institution Zone (IZ) record linked to the CZ record for local use. Updating the CZ records themselves is problematic for individual institutions because Ex Libris maintains the CZ records with metadata supplied by the various vendors. Institutions do have the ability to add certain local fields to the CZ record, which creates the local fields in the IZ linked record. The local field must include special coding (local extension) to prevent the call number from being overwritten when Ex Libris updates metadata in the CZ. The team elected to use the 090 field with a local extension to insert call numbers into the IZ copies of the CZ records. Inserting 60,000 LC call numbers manually was not feasible. The Ebook Call Number Team experimented with two approaches to batch-processing LC call numbers into Alma. One option was to create a set and apply normalization and indication rules to insert the same call number via a job. A second option involved retrieving MARC records with the proper call number and overlaying them into Alma using an import profile. The team ultimately used both processes throughout the project. The second process (overlay) worked well during the automation phase, and the first process (normalization and indication rules) was useful during the cataloging quality-control phase.

In the third phase of the action plan, the team needed to devise a method to automate call number retrieval. Processes explored included using a macro to scrape call numbers from WorldCat, using OCLC Classify, the MarcEdit/Alma API, OpenRefine, the OCLC Connexion API, or ChatGPT. The macro process worked well, but it retrieved only the call number, not the whole MARC record. OCLC Classify created a call number from the OCLC subject heading, but OCLC discontinued this plug-in. Although OpenRefine is excellent for cleaning metadata and provides APIs for data retrieval, the team was unable to retrieve OCLC call numbers in batch. MarcEdit, an excellent cataloging tool, has an API for retrieving records from Alma, which could then be updated in MarcEdit and returned to Alma. Because MarcEdit's call number plug-in was tied to the discontinued OCLC Classify, the team could not use this tool.

The Connexion API had the greatest potential for the call number project because it enables OCLC searches that return either specific fields or complete MARC records. With assistance from the C&D Department's programmer, the team used this feature to retrieve call numbers, insert the local 090 extensions, and then reload the records into Alma. Because the USF Libraries are an OCLC member, it is permissible to retrieve OCLC records and call numbers via API and use them in the IZ of Alma. Institutions without a programmer on staff may find this solution difficult because it requires extensive programming expertise to write API data retrieval code. The team lastly explored using ChatGPT to

retrieve call numbers. Although ChatGPT had great potential for future projects, the team ultimately decided not to use it for this project due to the extensive training required to teach ChatGPT the LCC system.

The Metadata/Cataloging Unit collected statistics on the Ebook Call Number Project throughout the project to assess its success, evaluate its effectiveness, and inform improvements to the workflow for future endeavors. Each team member recorded consistent, identifiable metrics for each task and for every step within that task. Statistics collected at the beginning of the project indicated the number of ebooks in Alma with call numbers and which collections required remediation. Statistics collected during the project indicated the number of call numbers successfully corrected by the automation process, the number requiring manual updates, and the number requiring manual correction by catalogers. Statistics collected at the end of the project determined the total number of ebooks that now had an LC call number and the number that still required one. These statistics proved invaluable to the team for monitoring progress and assessing the success of the automation process. Figure 5 shows the project's statistics tracking spreadsheet.

Implementation Workflow

Once the action plan phases were completed, the gathered information was used to develop an implementation workflow for this project. The workflow consists of the following processes: record set creation, automation, quality control, and report creation.

The first process in the implementation workflow involved creating MARC record sets for ebooks that required call numbers. A team member created sets in Alma for each electronic ebook collection activated in the CZ. Each set was then exported as full MARC bibliographic records. These MARC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	# of records with call number automatically inserted	# of records with call number manually inserted	# of records with Automation Correction	# of records with 050 inserted post run	# of records NOT imported	Ready for Liang	load back to Alma	Marc records no 050 field (after removing collection marc)	Vendors	Alma Electronic Collection Name	Current Number of Portfolios Active in IZ	MARC Total Count(from MarcEdit)	Call Number Percentage 050/total	Notes
27	43	0	6			X	X	43	Bloomsbury	Bloomsbury Medieval Studies	199			
28	5	0	0			X	X	5	Bloomsbury	Bloomsbury Medieval Studies 2020-2021 Update	24			
29	4	0	0			X	X	4	Bloomsbury	Bloomsbury Medieval Studies Annual Update 2022	37			
30	3	0	0			X	X	3	Bloomsbury	Bloomsbury Medieval Studies: Core Collection - Annual Update 2023	16			
31	67	0	0			X	X	67	Bloomsbury	Bloomsbury History: Theory & Method - Launch Collection	219			
32	14	1	0			X	X	15	Bloomsbury	Manchester Medieval Sources Online (MMSO)	41			
33	802	4	802			X	X	806	Bloomsbury	Bloomsbury Drama Online Core Collection US/Canada	1167			
34	13	0	13			X	X	13	Bloomsbury	Bloomsbury Drama Online Core Collection 17-18 Update US/Canada	116			
35	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online: Core Collection 2018-2019 Update US	98			
36	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online Core Collection - 2020 Annual Update US	54			
37	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online: Core Collection Annual Update 2021	63			

Figure 5. Statistics tracking for call number insertion.

records were run through MarcEdit, and field counts for all call number fields were recorded; these were later used to assess the effectiveness of the automation process. The team stored this data in a project statistics spreadsheet organized by vendor and ebook collection for easy retrieval and analysis.

The second process in the workflow involved the automation process. The programmer created a program that retrieved LC call numbers from OCLC by using the Connexion API. These call numbers were inserted into a 090 local extension field. The metadata specialist created an import profile to reload MARC records into Alma with the local 090 extension field. The MARC records were then reloaded into Alma by vendor and collection. The total number of automated call numbers was recorded on the project statistics spreadsheet.

The third process of the workflow involved quality control. While the Ebook Call Number Team felt automation was the best approach, for quality-control purposes, catalogers still needed to monitor it. The complex copy catalogers checked the reloaded records for accuracy and corrected any errors in the automation. The catalogers also manually inserted call numbers into records that could not be automated. The team performed final quality control on project completion and recorded the statistics in the project statistics spreadsheet.

The final process in the workflow involved creating analytical queries for building report cards. The metadata quality-control manager used Alma Analytics to build custom reports for the C&D Department, which were in turn used to create USF Libraries report cards.

Record Set Creation Process

A team member conducted an advanced search in Alma to ensure the data was relevant and accurately reflected the intended collection. The search sought book titles in electronic collections that lacked call numbers. Figure 6 displays search parameters used in Alma.

The results were saved as a set, and then the bibliographic data from the set was exported using Alma's "Export Bibliographic Records." The downloads were too large to handle efficiently, so the files were

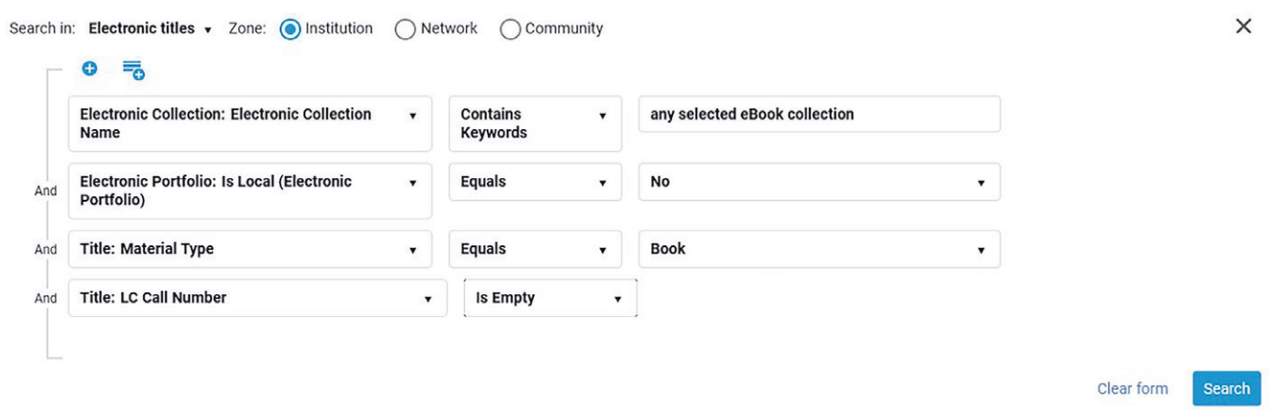


Figure 6. Advanced search in record set creation process.

parsed into smaller sets of 1,000 records. This approach improved processing times and allowed the programmer to detect errors more easily. Renaming the parsed files to a standard naming convention helped avoid issues during automated processing.

Automation Process

This section provides a comprehensive automation workflow for the Ebook Call Number Project. The automation process used Python to enhance bibliographic records. The heart of the automation process was a Python program designed to streamline the retrieval and integration of call numbers into MARC records.

The core of the automation strategy was a custom Python program.[18] Inspired by earlier efforts to use a macro for automation, the programmer developed a solution tailored to the project's specific needs. In the article "Opening the Door: A First Look at the OCLC WorldCat Metadata API," Terry Reese mentioned that OCLC WorldCat APIs support read and write access to the WorldCat database, as well as offer new functions and workflows that directly interact with the WorldCat database.[19] Using the capability described by Reese, the programmer wrote a Python program that used an API to search for bibliographic records one by one in the OCLC database, using different identifiers and prioritizing them based on their relevance and availability. Upon retrieving the call numbers from bibliographic records, the program automatically inserted them into the MARC records, specifically into the 090 field. This automation ensured that bibliographic records were accurately updated with the correct call numbers, requiring minimal manual intervention.

Before writing the Python program, the programmer needed to complete two tasks. The first one was to request a WorldShare Key at the WSKey Management platform. Obtaining this key was crucial for integrating OCLC's resources into the automation process because it provided access to two critical services: the WorldCat Search API, which enabled searching for bibliographic records in WorldCat; and the WorldCat Metadata API, which facilitated the retrieval and management of metadata from WorldCat. The second task was to prepare the programming tools. For this project, the programmer utilized PyCharm, specifically the professional version. PyCharm is a robust Integrated Development Environment (IDE) that supports Python programming with features such as code completion, debugging, and project management.[20] The program relied on several Python libraries to function effectively: Pandas 3.0.3, Pymarc 5.1.0, and BookOps-WorldCat 0.5.0. Python 3.9.8 was used to ensure compatibility with the libraries and tools.

The programming logic addressed the limitations of the OCLC API and ensured accurate retrieval of bibliographic information. The OCLC API could retrieve full bibliographic records with call numbers only when searching by OCLC number. If searches used other identifiers such as ISBN or titles, the API would return brief records containing OCLC numbers but not call numbers. The program's approach was to find the OCLC number and use it to retrieve the full MARC record and extract the first call number field found on the record.

Figure 7 displays a flowchart of the Python program. The following explains how the program processed a single record. First, the program read a MARC record and initiated a search using available identifiers. If an OCLC number was present, the program issued a follow-up request using that number to retrieve a full bibliographic record that included the call number. Once the call number was retrieved, it was automatically inserted into a local 090 field of the MARC record. If there was no OCLC number, then the program searched using other available identifiers and then retrieved brief records with OCLC numbers. The program then made a follow-up request using the OCLC numbers to fetch full bib records, which would include call numbers. Once the first call number was retrieved, it was automatically inserted into the local 090 field of the MARC record.

The Python program generated three distinct file types for each set of bibliographic records, each serving a specific purpose in the bibliographic data enhancement workflow. The program's primary output was a set of MARC records with call numbers automatically inserted into the local 090 field. These records overlayed existing records in Alma. The second output was a Comma-Separated Values (CSV) file that catalogers used for quality control, to review and correct, to assign call numbers as needed, or to make necessary adjustments. The third output consisted of program log files, which the programmer used to monitor the program's performance, check for correct execution, and identify any errors or issues.

According to Toaza and Esztergár-Kiss, the development and implementation of Python programs have significantly streamlined the bibliographic data management process.[21] By automating the retrieval and integration of call numbers using Python programming, the team enhanced accuracy, reduced manual effort, and improved overall efficiency.

The final step in the automation process was to overlay the MARC records in Alma with the primary output from the Python program. This was done using an import profile. The import profile was configured to ensure that records were updated correctly and that the new call numbers were properly incorporated into the existing bibliographic data. To optimize the import process and improve efficiency and the accuracy of statistical tracking, the 1,000-record files were combined by vendor when possible.

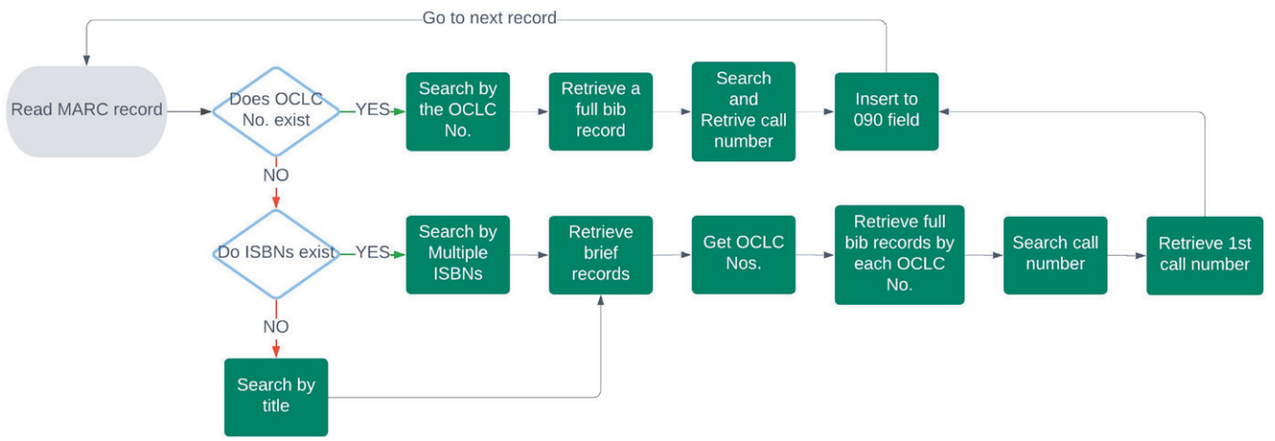
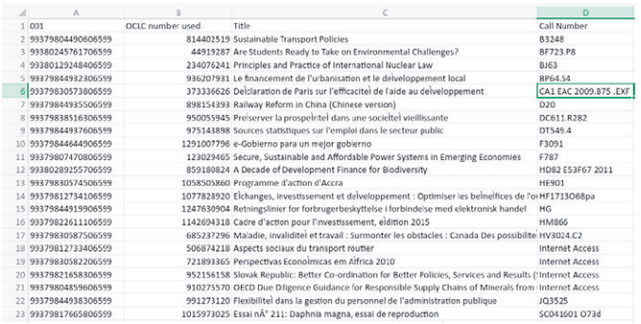


Figure 7. Customized Python program automation workflow chart.

Quality-Control Process

The third process in the Ebook Call Number Project workflow was quality control. Park and Tosaka indicate that it is essential to perform thorough quality control to ensure the accuracy of the bibliographic data.[22] The Python program, although efficient, occasionally retrieved incorrect call numbers or included irrelevant content. Catalogers manually reviewed the automated call numbers to verify that the assigned LC class was correct and relevant by matching the assigned call number to the subject headings; they corrected any discrepancies or errors identified during the review process directly in Alma. The catalogers also created call numbers for records that could not be automated.

Catalogers received reports in the form of CSV files as part of the Python program's second output. See figure 8 for an example of the fields in the quality-control report. After converting the CSV file to Excel, the catalogers verified that the retrieved call numbers were accurate.



	A	B	C	D
1	001	OCLC number used	Title	Call Number
2	9937980449060659	814402519	Sustainable Transport Policies	B3248
3	9938024576170659	44919287	Are Students Ready to Take on Environmental Challenges?	BF723.P8
4	9938012924840659	234076241	Principles and Practice of International Nuclear Law	BJ63
5	9937984493280659	936207931	Le financement de l'urbanisation et le dévelopgement local	BP64.34
6	9937983057380659	373336626	Déclaration de Paris sur l'efficacitél de l'aide au dévelopgement	CA1 EAC 2009.875 .EXF
7	9937984493550659	898154393	Railway Reform in China (Chinese version)	D20
8	9937983851630659	950055945	Préiserver la prospelrité dans une societé vieillissante	DC611.R282
9	9937984493760659	975143898	Sources statistiques sur l'emploi dans le secteur public	DT549.4
10	9937984464906659	1291007796	e-Gobierno para un mejor gobierno	F3091
11	9937980747080659	123029465	Secure, Sustainable and Affordable Power Systems in Emerging Economies	F787
12	9938028955706659	859180824	A Decade of Development Finance for Biodiversity	HD82 E53F67 2011
13	9937983057450659	1058505860	Programme d'action d'Accra	HE901
14	9937981273410659	1077828920	Elchanges, investissement et dévleoppement : Optimiser les belnelfices de l'oi	HF1713O68pa
15	9937984491990659	1247630904	Retningslinier for forbrugerbeskyttelse i forbindelse med elektronisk handel	HG
16	9937982611060659	1142694318	Cadre d'action pour l'investissement, éldition 2015	HM866
17	9937983058750659	685237296	Maladie, invaliditél et travail : Surmonter les obstacles : Canada Des possibilte	HV3024.C2
18	9937981273340659	506874218	Aspects sociaux du transport routier	Internet Access
19	9937983058220659	721893365	Perspectivas Econolmicas em Afrlica 2010	Internet Access
20	9937982165830659	952156158	Slovak Republic: Better Co-ordination for Better Policies, Services and Results (	Internet Access
21	9937980485960659	910275570	OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from	Internet Access
22	9937984493830659	991273120	Flexibilitél dans la gestion du personnel de l'administration publique	JQ3525
23	9937981766580659	1015973025	Essai nÂ° 211: Daphnia magna, essai de reproduction	SCD41601 O73d

Figure 8. Cataloger's quality-control report.

The cataloger used the report to identify and correct incorrect call numbers and to assign call numbers to items that had not received one. Although the percentage of incorrect call numbers was low, some were present. For example, if the wrong OCLC number was in Alma's CZ record, the program pulled call numbers from records that did not match the item. When no OCLC number was available, the program performed a title search, which sometimes yielded inaccurate results. Sometimes the OCLC record did not have an actual LC call number but contained phrases such as "Internet Access" or "Electronic Resource" in a call number field. Additionally, the call number field sometimes contained incomplete call numbers: it included only the alphabetical portion of the LC call number or call number ranges (e.g., AB12-56).

Depending on the number of call numbers that needed correction, the cataloger either intervened manually or ran a batch process. During manual correction, catalogers would search OCLC by title or ISBN to obtain a call number. If no record with an appropriate call number was found, the cataloger would search for subject headings in OCLC and/or the Alma CZ and use ClassWeb to determine a corresponding call number. Catalogers also used Dewey Decimal numbers (if present) and searched in ClassWeb to find an LC call number. As a last resort, the ebook itself was consulted to assign a call number for the item using subject analysis. Once a call number was identified, the cataloger would insert a 090 local extension field into the IZ record in Alma and delete any incorrect call number fields added through automation.

Typically, when more than 100 call numbers needed to be corrected or added, a batch process was used to save time. During a batch process, records with similar subject headings were assigned the same LC call number and updated in bulk using a normalization rule. Catalogers used the same procedure for determining batch call numbers as the earlier process for creating manual call numbers. Normalization

rules are queries used to modify bibliographic records in Alma. After creating a set of records in Alma, the cataloger would export the bibliographic records from Alma and use a MarcEdit tool to convert to a spreadsheet with the following fields: identifier, OCLC numbers, title, call number, and subject headings. Abrahamse wrote, "This tool can provide a useful overview of large or even not-so-large files of MARC records."[23]

The catalogers sorted the MarcEdit report by the MARC 050 field and removed records that already contained call numbers. These call numbers were most likely added to the IZ record after the team exported the original MARC record set. Next to be sorted are the local 090 fields that were incorrectly assigned by the automation process; these items were collected and converted into a set in Alma. Once the set was created, the local 090 fields were bulk deleted from the records using a normalization rule. The catalogers would then sort by subject headings, and main class call numbers were assigned based on these headings. It was not necessary to create cutters for the call numbers because the USF Libraries report cards represent twelve broad discipline areas and use only the top level of the LCC structure. This made it possible to add one general call number to the entire subject area. For example, items with a subject heading about soil moisture were assigned the LC call number S594. All items assigned the call number S594 were grouped into a single set in Alma. Before running the job that inserts the correct local 090 fields, the cataloger would edit the normalization rule associated with the job. Using the same example, the normalization rule was adjusted to insert the value S594 into the local 090 field. The normalization rule must be updated to reflect the call number to be inserted into the records before each run. Examples of the normalization rules created for this project are in figure 9.

Analytics Reports Process

With the data now enhanced in Alma through the completion of the Ebook Call Number Project and the uploading of COUNTER reports via SUSHI harvesting, the Report Card process could now be automated. The fourth process in the workflow was to build Alma Analytics reports to facilitate the creation of the USF Libraries report cards. LCC was built into Alma's reporting capabilities. This allowed the construction of analytics reports to capture inventory and usage counts for print and electronic resources by LC Class. The consortium built Alma Analytics reports by LCC for USF Libraries; the Ebook Call Number Team further customized these reports by adding LCC to the consortia-created IPEDS reports. This allowed USF Libraries to report on the print book and journal inventory counts, as well as ebook and e-journal counts, by discipline for the annual report cards. The usage statistics

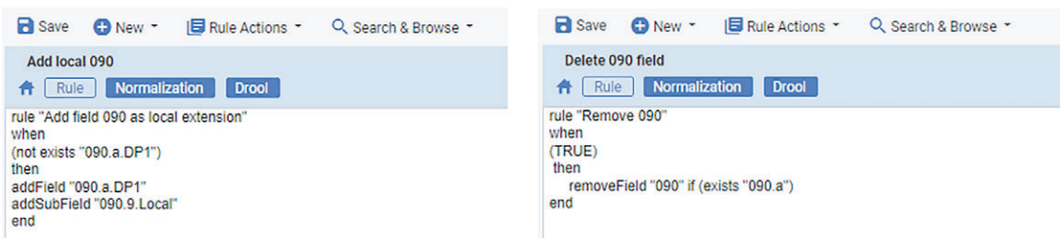


Figure 9. Examples of Alma normalization rules.

were compiled into two separate analytics reports: one for print loan usage by the LC class and one for electronic COUNTER report usage by the LC Class.

To test the reliability of the Alma Analytics–created report cards, the previous year's data from the manual report card creation was compared with the Alma Analytics–created reports of the current year; the number counts for inventory and usage were found to be very similar. The Alma Analytics report card data was also compared with vendor data and SUSHI reports to verify accuracy. This new automated process using Alma provided much more comprehensive data and generated incredible time savings in report creation.

Outcomes

The Call Number Project Team kept statistics on the various components of the project throughout the entire process. Initially, statistics were collected on the current percentage of ebooks with call numbers. During the project, the team collected statistics on the following: the number of call numbers added by automation; the number of automated call numbers requiring corrections; the number of call numbers manually added by catalogers; and the number of records with 050 fields added since the start of the project. Finally, toward the end of the project, the team ran analytics reports to determine how many ebooks had call numbers after the project and performed advanced searches in Alma to identify ebooks without call numbers for future cleanup.

Considering the ratio of call number records before and after the project, the data confirms that before the project, there were 631,129 records with call numbers and 59,040 without call numbers. After the project, there were 685,724 records with call numbers and only 4,445 without call numbers. All records included in the project results exclude primary resources, music scores, and audiovisual electronic materials because the focus of the project was ebooks. The 4,445 records without call numbers may either no longer exist or were unintentionally excluded; these records will be reconciled in future cleanup efforts. Looking at the same statistics as percentages, before the project, 90 percent of the ebook records contained call numbers, and 10 percent lacked them. After the project was complete, 99 percent of the ebook records contained call numbers, and only 1 percent lacked them.

Considering the time and labor required to insert call numbers into the local 090 field, the data indicates that of the 59,040 records with no call numbers, 44,653 call numbers were inserted through the automation process, and only 2,641 of the automated call numbers required correction. Finally, the catalogers had to enter 9,942 call numbers manually. The automation data, shown as percentages, indicates that 94 percent of automated call numbers were correct, and only 6 percent required modification. Data further shows that 78 percent of the call number insertions were automated, 5 percent needed corrections as a result of the automation process, and 17 percent required catalogers to manually insert call numbers. See figure 10 for automation statistics.

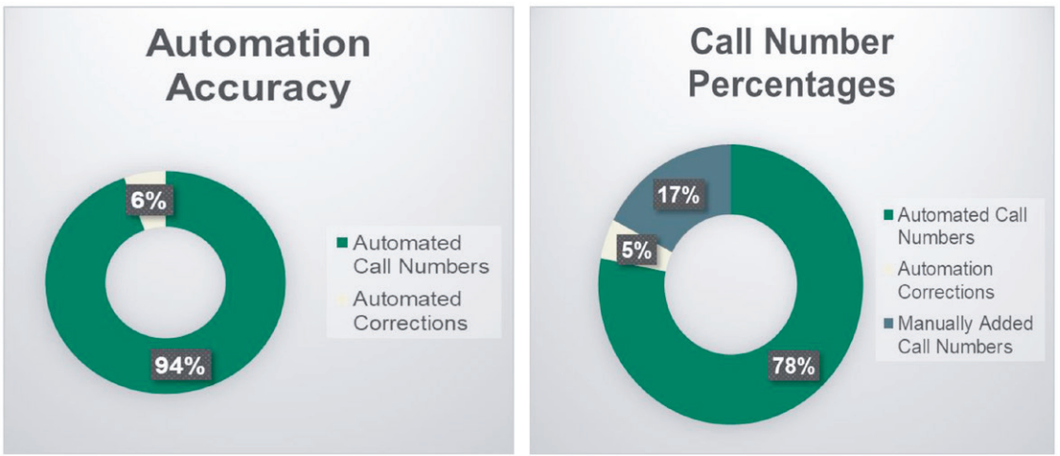


Figure 10. Statistics on the call number automation.

Conclusions

The automation process retrieved call numbers with an accuracy rate of greater than 94 percent. Not only was the Ebook Call Number Project successful, but its benefits were numerous and immediately apparent. Statistics for the annual report cards no longer had to be compiled manually from vendor websites, reducing the time required to produce the report cards. The accuracy of Alma Analytics reports improved by almost 50 percent, making the report cards more accurate. Adding LCC numbers enabled the USF Libraries to sort library resources by discipline, a key factor in accreditation reports for new programs. USF Libraries were now able to provide more comprehensive report cards to the various departments and colleges at the USF.

Lessons Learned

The Call Number Project Team identified several valuable lessons; however, five key lessons may help other institutions undertaking this project. First, due to the nature of Alma, most of the USF Libraries' electronic records are linked to the CZ. Still, some ebook collections are loaded locally and are not available in the CZ. The team did not consider this possibility when running sets at the beginning of the project and inadvertently captured some local records that were not linked to the CZ during the reload process. The import profile was set to update the CZ records with a local extension that placed the call number in our IZ linked records; accordingly, these local records would not overlay because they were not linked to the CZ and needed to be updated manually. The typical error was "Match Found but Not Imported." As noted by Chen, Kim, and Montgomery, overlaying local bibliographic records often requires specific import profiles and workflows.[24] In the future, these local records would be excluded from the record set and updated separately using a different import profile. Second, the team learned that between the time the records were exported from Alma and reloaded, some records had been removed because of CZ updates and automatic holdings updates from certain vendors. Because these records were no longer in Alma, they did not reload. The typical error message was "No Match Found."

This was also an issue noted by Rathemacher, Ragucci, and Doellinger in their article "Don't Wait, Automate! Industry Perspectives on KBART Holdings Automation."[25]

Third, the team discovered that exporting an entire collection from Alma sometimes included records for materials other than ebooks. Some collections in the CZ include ebooks, audio, streaming media, and e-journals in the same collection, all for the same vendor and on the same platform. This oversight highlighted the importance of setting correct facets from the beginning. In the future, the team will need to filter collections by material type and export only ebook records. Fourth, the team discovered that ebooks purchased by the consortium were activated in the Network Zone for distribution to consortia members; USF Libraries were unable to automate updates in batch to Network Zone records. Finally, during the MARC record export process in Alma, a collection level bib record was exported in each file. This affected both the automation process and the project's statistical analysis. To resolve this, a team member used MarcEdit to remove these collection bib records before finalizing the files.

Although the Python program has proved effective, the API request limitations have been a challenge. One potential issue was the risk of being temporarily blocked by OCLC because of excessive API requests. Several times, the programmer encountered temporary access restrictions when the program sent numerous requests in a short period. The typical resolution was to wait a couple of hours before resuming the requests. To mitigate this risk, the team opted to keep files small, limiting each file to 1,000 records. This approach, recommended by Terry Reese, helped manage API load and reduced the likelihood of hitting request limits.[26]

The team also established best practices to ensure quality control. For example, when searching for records in OCLC, if there was no OCLC number or ISBN, the Python program would perform a title search for the record in OCLC. Although this had a 50 percent success rate, it also returned incorrect call numbers that required manual correction. Future projects will not include title searching; records with neither an OCLC number nor an ISBN will be moved directly to manual processing.

Many of the corrections to automated changes were due to faulty metadata in the MARC records. Many OCLC records contained "electronic resource" or "online" in the call number field instead of an actual call number, and these records had to be updated manually. The programmer resolved this issue by programmatically filtering these exceptions from the results and then searching for another record with an actual call number. Another change for future projects will be to include subject headings in the program reports for each record. Without the subject headings, catalogers matched call numbers based on the title alone, which could introduce a margin of error depending on the depth of context the title conveyed. Including subject headings from the project's outset will enable more accurate verification of call numbers and eliminate the need to run sets through MarcEdit to find them.

Overall, the workflow for the Ebook Call Number Project performed well, with only a few minor, unexpected issues that the team corrected as the project progressed. Some of the workflow improvements include filtering Alma searches by material type and filtering out local collections in Alma. A separate workflow will be designed to add call numbers to local collections, including

creating an import profile to update the call numbers of local records and filtering local collections only when searching. Removing title searching from the programming improved the API's accuracy, and writing advanced scripting filters eliminated false hits in the call number field, further improving the automation process. During manual call number insertion, adding subject heading retrieval to the automation process made manual cataloging of call numbers easier for the cataloger. These improvements have made a big difference in the workflow performance.

Alternative Methods for Institutions Without a Programmer

The workflow discussed in this article was created through an iterative process. During the development phase, the Call Number Project Team explored various alternative techniques to accomplish different parts of the project. These were ultimately discarded in favor of a more efficient process developed by a programmer. It's important to note that not all institutions have programming resources available. Therefore, the authors document alternative methods to provide alternative paths to accomplish a similar project. Without programming expertise, UI.Vision's RPA, a macro-recording program (https://ui.vision/), and ChatGPT (https://chatgpt.com/) are two tools that librarians could consider using to automate the process of finding call numbers for ebooks, with several caveats discussed in the following paragraphs.

UI.Vision RPA was successfully used to automate the process of looking up ISBNs on the FirstSearch website and retrieving LC call numbers. This free browser extension allows users to record macros, which means that mouse clicks and keystrokes can be recorded as the user navigates through a website. The strengths of UI.Vision RPA include user-friendly basic functionality, the ability to use spreadsheets for input and output, thorough product documentation (https://ui.vision/rpa/docs), and free cost.

The completed workflow involved providing a spreadsheet of ISBNs to a macro, which then searched the OCLC FirstSearch website and recorded the LC call number back onto the spreadsheet. Although successful in trials, the drawback was the time required to obtain results. The macro searched the website in real time, meaning it had to fill out forms, click buttons, and load webpages for each search. Although effective, this process was inefficient compared with a programming solution. Developing the macro was also not as straightforward as recording. The data returned on search result pages was not static and not always in the same location. Extracting the LC number required identifying the correct Xpath, and the macro also needed additional adjustments to account for variables such as page loading times.

Macros created using UI.Vision RPA can automate the process of retrieving call numbers from FirstSearch. This tool could be beneficial for institutions that lack programming support; however, setting up macros can be complex and finicky. Fortunately, UI.Vision provides excellent documentation, and the community offers strong support (https://forum.ui.vision/).

ChatGPT was used experimentally to identify call numbers for a fringe case. This project was completed in ChatGPT-4, the current version at the time, using the Data Analyst GPT, a paid service. Since then, the results have been replicated in the free version of ChatGPT-5. One collection of ebooks that needed

LC numbers lacked ISBNs but had subject headings. ChatGPT was used to look up the subject headings from the LC Subject Headings Word document and return the LC number.

The advantage of using ChatGPT in this case was that it could search across different file types to return results. ChatGPT was given an Excel file containing the ebook titles and a Microsoft Word file with the LC Subject Headings. It was able to look up the subject headings from the Excel file in the Word file and then print the results back into the Excel file. ChatGPT's strengths include its ability to handle multiple file types and to perform complex tasks via conversational prompts without requiring programming or specialized knowledge. However, obtaining the exact output often requires iterative experimentation with the application. A careful examination of the results is also required because ChatGPT can make errors or be incomplete in its responses.

The use of ChatGPT in cataloging has been controversial. For instance, the article "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging" asserted that ChatGPT could make successful MARC records.[27] This set off a flurry of letters to the editor and a scathing review of the article titled "ChatGPT Not Useful as a Tool to Streamline Library Cataloguing Processes."[28] The criticisms centered around how error-ridden the artificial intelligence (AI)-generated MARC records were and how the author, not a cataloger, was unqualified and unable to determine the quality of the records.[29] Lowagie, in his careful, nuanced approach to creating an ethical universal catalog, says that AI, as a tool, "plays a pivotal role in recognizing patterns and relationships across datasets."[30] This project takes a similar view of ChatGPT as a singular, useful tool within a broader methodology. In this use case, ChatGPT accomplishes a specific task with easily verifiable results, making it a useful tool in a cataloger's arsenal.

Automating the retrieval of call numbers for ebooks is still possible even without programming resources. Applications such as UI.Vision RPA and ChatGPT can reduce substantial burdens by automating more labor-intensive tasks in a project, improving reporting and visibility of e-resources.

Future Plans

On completion of the project, USF Libraries' immediate plans included enhancing metadata for e-journals; however, lessons learned from the ebook enhancement enabled the e-journal project to proceed more quickly, and it was completed in the summer of 2024. The Metadata/Cataloging Unit plans to run an annual update of ebooks and e-journals using a modified workflow. Because annual updates will be in smaller batches, using AI applications to validate call numbers against subject headings and enhance metadata could improve the automation process. Additionally, the unit is considering expanding the metadata enhancement to include subject headings and call numbers. To make reports and statistics available to stakeholders across USF Libraries, the Metadata/Cataloging unit will create a dashboard that allows any department to run reports as needed. High-quality metadata is critical to improving the library's ability to generate accurate reports that display library resources for the benefit of the university.

Notes

1. Association of Research Libraries. Systems and Procedures Exchange Center, *Qualitative Collection Analysis: The Conspectus Methodology* [SPEC Kit,0160-3582 ;151] (Association of Research Libraries, Office of Management Services, 1989). https://catalog.hathitrust.org/Record/001078642.
2. Jennifer Benedetto Beals and Ron Gilmour, "Assessing Collections Using Brief Tests and WorldCat Collection Analysis," *Collection Building* 26, no. 4 (2007): 104–7, https://doi.org/10.1108/01604950710831898; Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *College & Research Libraries* 67, no. 1 (2006): 35–49, https://doi.org/10.5860/crl.67.1.35.
3. Michael Luther and Ana Guimaraes, "Applying the Principles of Total Library Assessment to Inform Sustainable Collection Development," *Proceedings of the 2016 Library Assessment Conference* (2017): 181–87; Mary H. Munroe and Jennie E. Ver Steeg, "The Decision-Making Process in Conspectus Evaluation of Collectons: The Quest for Certainty," *The Library Quarterly (Chicago)* 74, no. 2 (2004): 181–205, https://doi.org/10.1086/382846.
4. Jennifer Benedetto Beals, "Assessing Library Collections Using Brief Test Methodology," *Electronic Journal of Academic & Special Librarianship* 7, no. 3 (2006): 1, https://research.ebsco.com/linkprocessor/plink?id=2aae0fc0-a059-30d2-aa1f-e827f0d84517.
5. Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *College & Research Libraries* 67, no. 1 (2006): 35–49, https://doi.org/10.5860/crl.67.1.35; Mary H. Munroe and Jennie E. Ver Steeg, "The Decision-Making Process in Conspectus Evaluation of Collections: The Quest for Certainty," *The Library Quarterly (Chicago)* 74, no. 2 (2004): 181–205, https://doi.org/10.1086/382846.
6. Karen C. Kohn, "Usage-Based Collection Evaluation with a Curricular Focus," *College & Research Libraries* 74, no. 1 (2013): 85–97, https://doi.org/10.5860/crl-295.
7. Sheila S. Intner and Elizabeth Futas, "The role and impact of Library of Congress Classification on the assessment of women's studies collections," *Library Acquisitions: Practice & Theory* 20, no. 3 (1996): 267–79, https://doi.org/10.1016/0364-6408(96)00067-1.
8. Jonathan Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus: Providing Subject Categorization for Users of Electronic Resources," *Journal of Internet Cataloging* 5, no. 4 (2002): 67–80, https://doi.org/10.1300/J141v05n04_05; Kimberly Looby, "Subject Librarians Vs Collection Data: How to Use Data to Visualize a Small Monograph Collection," *Technical Services Quarterly* 38, no. 1 (2021): 33–53, https://doi.org/10.1080/07317131.2020.1854576.
9. Stephen Paul Davis, "HILCC: A Hierarchical Interface to Library of Congress Classification," *Journal of Internet Cataloging* 5, no. 4 (2002): 19–49, https://doi.org/10.1300/J141v05n04_03.
10. Karen Harker, Janette Klein, and J. Laurel Crawford. "Multiplying by Division: Mapping the Collection at University of North Texas Libraries," *Where Do We Go From Here?: Charleston Conference Proceedings 2015* (2016): 333–40, https://doi.org/10.5703/1288284316279.
11. Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus," 67–80; Looby, "Subject Librarians vs Collection Data," 33–53.
12. Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus," 67–80.

13. Intner and Futas, "The role and impact of Library of Congress Classification on the assessment of women's studies collections," 267–79.

14. Michael Hughes, "How I Stopped Worrying and Learned to Love the Usage Data," *College & Research Libraries* 81, no. 7 (2020): 1168–83, https://doi.org/10.5860/crl.81.7.1168; Michelle Leonard and Steven Carrico, "Developing a Sustainable Collection Assessment Strategy," *Proceedings of the 2016 Library Assessment Conference* (2017): 172–80.

15. Suzanne Saskia Kemperman et al., "Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper" (OCLC Online Computer Library Center Inc., 2014): 2–24, https://www.oclc.org/content/dam/oclc/reports/data-quality/215233-SuccessStrategies.pdf.

16. Isidoro Gil-Leiva, Maria Carolina Andrade e Cruz, Franciele Marques Redigolo, and Mariângela Sptti Lopes Fujita, "Assignment of Subject Headings and Classification Codes to e-books in South American University Libraries," *Revista Española de Documentación Científica* 47, no. 2 (2024): e385, https://doi.org/10.3989/redc.2024.1513.

17. Donna E. Frederick, "EBook Discovery Metadata," in *Managing E-Book Metadata in Academic Libraries: Taming the Tiger* (Chandos Publishing, 2016): 105–201.

18. USF Libraries Call Number Automation Python Programming is available on request. Please contact Bonita Pollock at pollockb1@usf.edu.

19. Terry Reese, "Opening the Door: A First Look at the OCLC WorldCat Metadata API," *Code4Lib Journal*, no. 25 (2014), https://journal.code4lib.org/articles/9863.

20. JetBrains, "PyCharm," accessed December 10, 2024, https://www.jetbrains.com/pycharm/.

21. Bladimir Toaza and Domokos Esztergár-Kiss, "Automated Bibliometric Data Generation in Python from a Bibliographic Database," *Software Impacts* 19 (2024): 100602, https://doi.org/10.1016/j.simpa.2023.100602.

22. Jung-Ran Park and Yuji Tosaka, "Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms," *Cataloging & Classification Quarterly* 48, no. 8 (2010): 696–715, https://doi.org/10.1080/01639374.2010.508711.

23. Ben Abrahamse, "A Cataloger's Toolkit, Part Two: Using MarcEdit with Spreadsheets," *Technicalities* 42, no. 6 (2022): 16–18.

24. Mingyu Chen, Misu Kim, and Debbie Montgomery, "Ebook Record Management at The University of Texas at Dallas," *Technical Services Quarterly* 33, no. 3 (2016): 251–67, https://doi.org/10.1080/07317131.2016.1169781.

25. Andrée Rathemacher, Matthew Ragucci, and Stephanie Doellinger, "Don't Wait, Automate! Industry Perspectives on KBART Holdings Automation," *The Serials Librarian* 82, no. 1–4 (2022): 91–7, https://doi.org/10.1080/0361526X.2022.2019545.

26. Reese, "Opening the Door: A First Look at the OCLC WorldCat Metadata API."

27. Richard Brzustowicz, "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging," *Information Technology and Libraries* 42, no. 3 (2023), https://doi.org/https://doi.org/10.5860/ital.v42i3.16295.

28. Andrea Miller-Nesbitt, “ChatGPT Not Useful as a Tool to Streamline Library Cataloguing Processes,” *Evidence Based Library and Information Practice* 19, no. 2 (2024): 145–7, https://doi.org/https://doi.org/10.18438/eblip30524.

29. Christine DeZelar-Tiedman, “Response to ‘From ChatGPT to CatGPT.’” *Information Technology and Libraries* 42, no. 4 (2023), https://doi.org/https://doi.org/10.5860/ital.v42i4.16991.

30. Hannes Lowagie, “From Bias to Transparency: Ethical Imperatives in AI-Based Library Cataloging” (paper presented at the IFLA WLIC 2023 Satellite Meeting on Universal Bibliographic Control, Brussels, Belgium, August 18, 2023), https://repository.ifla.org/items/351270b4-31b2-4481-a67f-8412e48198bb.