Library Resources & Technical Services

Core LEADERSHIP
INFRASTRUCTURE
FUTURES

65 | 4

Library Resources & Technical Services

ISSN 2159-9610 October 2021 Volume 65, No. 4

Cover image: Mary Beth Weber, Laurita Winery, New Egypt, New Jersey, June 6, 2021.

Editorial

Farewell, Friends

Mary Beth Weber

This is my last *LRTS* editorial. Things have certainly changed in the nine years that I have held this appointment. When my predecessor presided over her last meeting with the board, it was in person at the ALA Annual Conference. It was an emotional, and also celebratory, meeting. In contrast, my final meeting with the editorial board will be via Zoom and as Core members. Never would I have imagined how ALCTS would change, including the departures of former Executive Directors Charles Wilt and Keri Cascio. When I won the ALCTS slogan contest in 2017, I did not anticipate the merger of LITA, ALCTS, and LLAMA to become Core. I later participated in the planning of the Core Exchange in 2020, including serving as the Exchange's blog coordinator. Working with participants from the three divisions, and under the leadership of Kristin Martin, was an amazing experience and demonstrated our ability to collaborate to make Core strong and sustainable.

Technical services has also evolved, partly due to quarantine. No one could have anticipated that we would live and work in isolation for more than a year. Many areas of technical services continue to handle physical materials, including acquisitions, cataloging, collection development, and preservation. However, there has been a gradual shift to electronic and digital that was accelerated by COVID and library closures. Once this type of change has taken place, it is not practical and often not possible to return to the former way of doing things. Drastic circumstances frequently drive change.

What has not changed is the need to educate, innovate, and share. This is what we do as an association and profession. One way that this is done is by attending conferences and giving presentations, and through publishing research to benefit others. Consider submitting a paper to *LRTS*. Share your ideas and experiences. Contact me if you have an idea. If you are conducting a survey or study, the results and analysis may form the basis of a thought-provoking research paper.

I thank ALCTS for giving me so many opportunities to participate and contribute to the profession. I was fortunate to have served as the editor of the *ALCTS Newsletter Online* for six years, followed by my two terms as *LRTS* editor. Both roles brought me into contact with many people, both within ALA and outside our organization. Serving as *LRTS* editor has been an honor and privilege. My term as editor ends in December 2021, and the plan is to have a new editor in place by October 2021. Please consider applying for this position or nominating a colleague.

I would now like to turn your attention to this issue of *LRTS*.

- In "Name and Subject Heading Reconciliation to Linked Open Data Authorities using Virtual International Authority File and Library of Congress Linked Data Service APIs: A Case Study featuring Emblematica Online," Tang (Cindy) Tian, Timothy W. Cole, and Karen Yu discuss
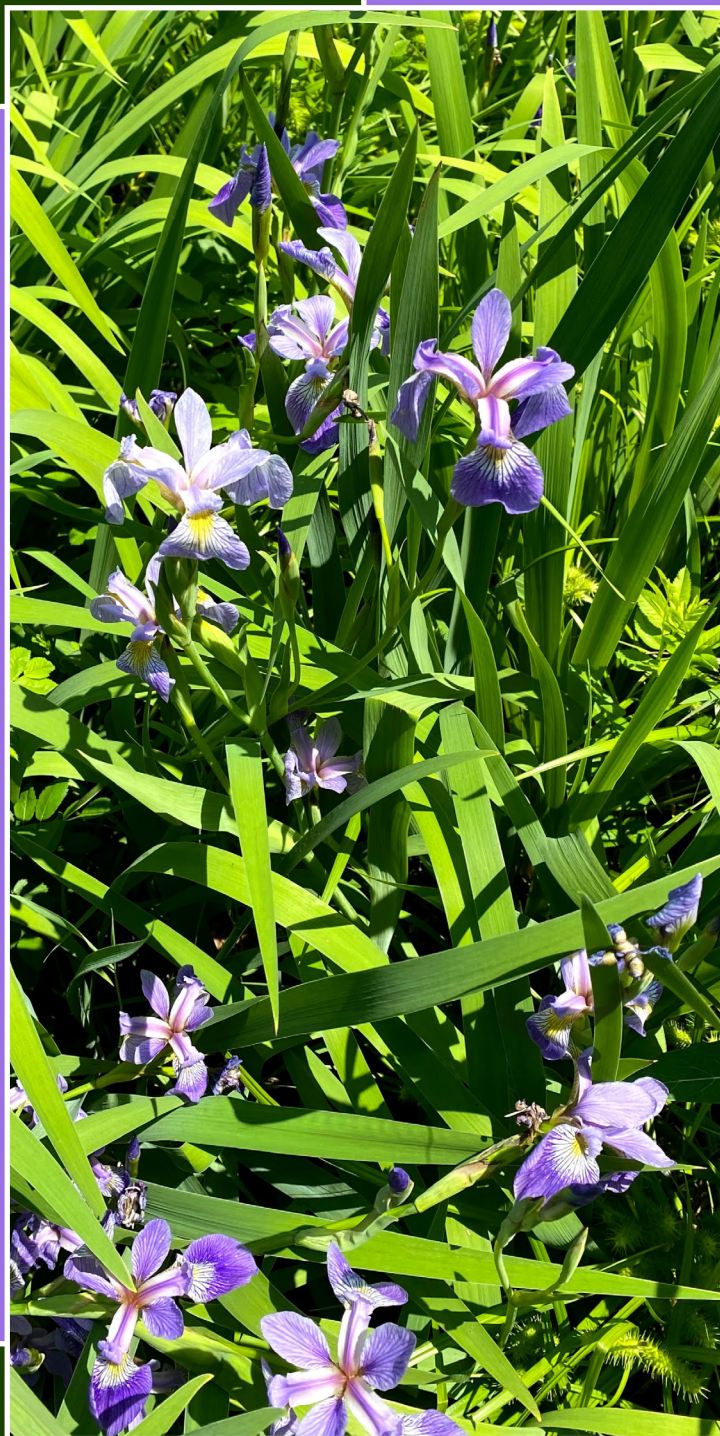
Emblematica Online to facilitate discovery of digitized emblem books. Emblematica Online incorporates linked open data in its design. The Virtual International Authority File and Library of Congress Linked Data Service APIs were used to reconcile name and subject strings from legacy catalog records with global authoritative links from LOD resources. Their case study reports on the automated reconciliation process used and examines the efficacy of the APIs to reconcile name and subject heading entities.

- "Dispersed Collections in Exile: Thai Collections in Libraries outside of Thailand," by Hollie White and Songphan Choemprayong, investigate how libraries outside Thailand manage Thai collections, particularly acquisitions and cataloging. The authors adopted Merton's Insiders/Outsiders doctrine and applied an exploratory mixed-method approach that used observation and interview methods.

- "Exploration of Subject Representation and Support of Linked Data in Recently Created Library Metadata: Examination of Most Widely Held WorldCat Bibliographic Records" presents the results of an examination of subject representation in the most recently created library metadata records. The study examined the level and patterns of application of subject controlled vocabularies. Vyacheslav Zavalin, Oksana L. Zavalina, and Shawne D. Miksa describe co-occurrences between various subject representation data elements and between subject controlled vocabularies within the records were explored.

- Book reviews courtesy of *LRTS* Book Review Editor Elyssa Gould.

In closing, I would like to recognize the following people and express my gratitude to them: Charles Wilt, Kerry Cascio, Brooke Morris-Chott, Tim Clifford, Elyssa Gould, Chelcie Rowell, Peggy Johnson, Miriam Palm, Norm Medeiros, Pamela Bluh, Dina Giambi, Susan Davis, Christopher Cronin, Jennifer Bazely, Ginger Williams, Oksana Zavalina, Karen Kiorpes, Hayley Moreno, and Kalan Knudson Davis. I also worked with many amazing authors during my term as *LRTS* editor and wonderful people in ALCTS, and now Core. And I am now contemplating the next phase of my career.

Name and Subject Heading Reconciliation to Linked Open Data Authorities using Virtual International Authority File and Library of Congress Linked Data Service APIs

A Case Study featuring Emblematica Online

Tang (Cindy) Tian, Timothy W. Cole, and Karen Yu

Libraries are actively exploring ways to use Linked Open Data (LOD) services to enhance discovery and facilitate the use of collections. Emblematica Online, which provides integrated discovery of digitized emblem books, incorporates LOD in its design. As an implementation prerequisite, the Virtual International Authority File (VIAF) and Library of Congress (LC) Linked Data Service APIs were used to reconcile name and subject strings from legacy catalog records with global authoritative links from LOD resources. This case study reports on the automated reconciliation process used and examines the efficacy of the APIs in reconciling name and subject heading entities. While a majority of strings were successfully reconciled, analysis suggests that data cleanup, rigorously consistent formatting of metadata strings, and addressing challenges in existing LOD resources and services could improve results for this corpus.

Tang (Cindy) Tian (ttian@nd.edu) is a Metadata Services Librarian at Kresge Law Library, University of Notre Dame. **Timothy W. Cole** (t-cole3@illinois.edu) (retired), Timothy W. Cole, Elaine and Allen Avner Professor Emeritus in Inter-disciplinary Research, University Library, University of Illinois at Urbana-Champaign. **Karen Yu** (karen4@uchicago .edu) is Head of East Asia Technical Services, Joseph Regenstein Library, University of Chicago.

Emblematica Online is a web-based digital library that describes and supports the discovery of 1,406 retrospectively digitized facsimiles of rare emblem books that contain more than 33,000 individual emblems from seven research institutions: the Herzog August Bibliothek in Germany (466 books); the University of Illinois at Urbana-Champaign in Urbana, Illinois (421 books); the Getty Research Institute in Los Angeles (248 books); Duke University in Durham, North Carolina (197 books); Glasgow University in Scotland (43 books),

Utrecht University in the Netherlands (30 books); and the Newberry Library in Chicago (1 book). Early Modern emblem books expressed complex ideas in a compact and compelling form. Melding text and images, emblems (see figure 1) typically feature a tripartite structure consisting of a brief motto in Latin or a European vernacular language (*inscriptio*), an enigmatic illustration (*pictura*), and a textual epigram (*subscriptio*).[1] The emblem is more than the sum of its individual parts, however; the *inscriptio*, *pictura*, and *subscriptio* work together to produce a greater meaning, the goal of which is to challenge the reader intellectually and stimulate new thought and knowledge. Emblem collections were commonly published as books, but they also pervaded the decorative arts and appeared in other contexts. Analyses of emblems help scholars to develop a fuller understanding of both sacred and secular art of the period. The emblem is a critical genre in the study of Renaissance and Baroque Europe, owing both to its wide geographic spread and to the window it opens on the attitudes of the period concerning nearly every aspect of life, ranging from religion and politics to war and peace. Emblems suggest the presence of an intentioned, sophisticated strategy for repurposing, reorganizing, and reading texts and images through a system of parallels and analogies that narrow meaning to impart new perspectives or ideas.

Inherently, by their nature and because emblems embody both a rhetorical structure and a process, they are ideally suited to digital presentation in a Linked Open Data (LOD) context that can reflect semantic patterns of associative thought. For this corpus, a LOD approach enhances descriptive precision and facilitates interoperability across multiple, disparate, and widely distributed emblem book collections, thereby opening new ways for emblem scholars

Figure 1. A tripartite emblem with inscriptio, pictura, and subscriptio.

to explore emblem literature. The LOD-based *Emblematica Online* portal makes emblems appearing in retrospectively digitized emblem books more visible to scholars in related disciplines, such as art historians, historians of Renaissance and Baroque cultures, comparative literary scholars, and other scholars who are interested in the wider relationship between literature and the visual arts, theories of representation, and iconography.

The original book-level and emblem-level metadata describing emblem book volumes and the individual emblems they contain were initially transformed by each participating library from local MARC records and local emblem-specific metadata records into records conforming to the Metadata Object Description Schema (MODS) and community-based emblem SPINE standard schemas, respectively.[2] Note that the development of the SPINE metadata structure standard was only one part of a larger effort toward a set of community metadata agreements for describing both emblem books and the individual emblems that they contain.[3] This work continues and to facilitate interoperability has included the adoption (guided by experience with the *Emblematica Online* portal and its precursors) of MODS usage guidelines and high level data content standards. To create the current incarnation of *Emblematica Online*, MODS and SPINE metadata records were harvested and normalized by scripts as needed. MODS/SPINE records are maintained in the portal backend as machine-readable XML.

LOD features and functionality have become an essential part of *Emblematica Online* to enhance discovery and research. The key point to enable these features is automated metadata reconciliation that maps bibliographic metadata from text strings to global Uniform Resource Identifiers (URIs) in LOD authorities (in this paper "LOD authorities" refers to LOD resources that can be used as substitutes for more traditional library authorities in the context of *Emblematica Online* and similar corpora). As part of the metadata reconciliation process for this project, a preexisting Python script for normalizing and managing MODS/SPINE metadata was adapted to integrate the reconciliation workflow and produce Resource Description Framework (RDF) graphs serialized as JavaScript Object Notation for Linked Data (JSON-LD), a way to store LOD in JSON format.[4] Names and subject headings in bibliographic records are two of the most representative metadata types that are suitable for exposure as LOD as there are more LOD on the web that provide contextual information around these classes of entities and include relevant relationships. An entity in the LOD sense of an entity-attribute-value model refers to who or what the authority value is about, as opposed to mere text strings in traditional authority control approaches.[5] Specifically, two tools are integrated in the script to query name and subject
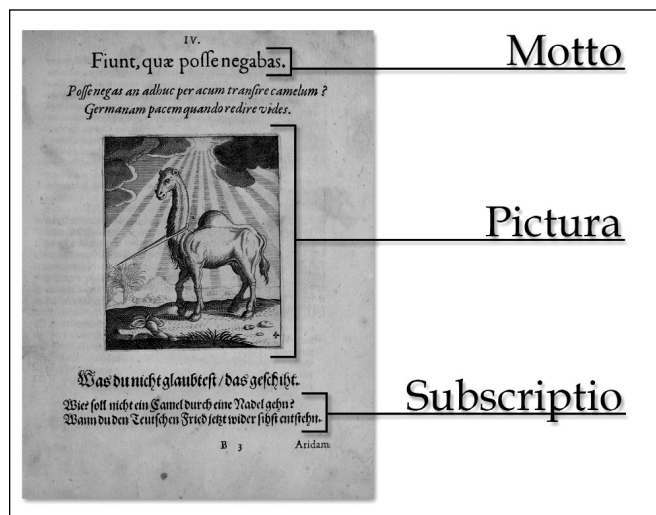
heading entities respectively: the Virtual International Authority File (VIAF) Auto Suggest API (hereafter VIAF Auto Suggest API) and Library of Congress (LC) Linked Data Service APIs (hereafter LC Linked Data APIs).[6]

Hosted by OCLC, VIAF is a name authority service that coalesces authority files of different, mostly national, library institutions from around the world. Successful reconciliation of name entities with VIAF authority records can enhance the user experience of digital library collections by accessing new and analytic information such as name variations for an author, titles associated with the author, and name forms in different languages. VIAF's Auto Suggest API automatically searches authority terms within VIAF based on a text passed in a query. LC's Linked Data Service provides base Uniform Resource Locators (URLs) with various search constraints to query LC ontologies and controlled vocabularies. This project uses the *aLabel* search constraint that will "only return a resource whose authoritative label exactly matches the searched term."[7] The end goal of using these two APIs is to enrich the original SPINE and MODS XML metadata with VIAF and LC authoritative links for name and subject heading entities. For subject heading entities, in addition to topical subject headings, the queries also consider genre and geographic subject headings as subject heading entities to reconcile.

This case study examines the reconciliation process in particular with a focus on two key issues:

1. Understanding the efficacy of the APIs used to reconcile name and subject heading entities;
2. To identify solutions to improve match results of digital collection metadata reconciliation to LOD authorities.

Literature Review

Application of LOD in library contexts is an active, current area of research. The application of LOD features to library collections and resources both increases the visibility of these resources on the web and provides end users with enhanced representations of primary sources, search results, and analytic information for research, especially within digital library special collections.[8] As application of LOD gains momentum in libraries, it is important to recognize the essential role of metadata reconciliation as part of planning and implementing LOD within libraries. According to the five-star scheme for evaluating the quality of LOD implementations, an implementation reaches the five-star level when the entities mentioned in a web application's data and descriptions (expressed in accord with RDF) are linked to other data sources and services on the Semantic Web.[9] For legacy data, e.g., bibliographic records

describing emblem books, this is achieved by data reconciliation, which supplements names and subject headings with URLs linking to additional, authoritative information about these entities. Proper communication between original or legacy metadata and appropriate LOD authorities provides interoperability and standardization for existing collections, along with matching a greater percentage of terms to existing controlled vocabularies.[10]

Research done as part of the initial implementation of LOD features in *Emblematica Online* identifies a few of MARC's limitations for use with RDF, especially in contrast to MODS and other metadata schemas.[11] The same research shows preliminary statistical findings for transforming MARC string-based authority control terms into VIAF and LC Subject Headings (LCSH) links. Related research in the context of *Emblematica Online* also includes an analysis of the XML-based SPINE metadata schema and the transformation of corpus metadata to more RDF compatible ontologies.[12] The findings from this earlier research demonstrate that to facilitate discovery and enhance the value to scholars of digitized emblem books, metadata must first be enriched with additional URIs and the workflow upgraded to normalize and transform existing emblem metadata, recognizing that the effort to do this would be substantial and needed to be fully worked out.[13] Since this research was published in 2017, subsequent work has been done to create Python scripts to automatically identify select entities in legacy metadata that could be enriched with authoritative links to LOD resources. This study was motivated by a need to report on the automated reconciliation process and examine the match rates of a subset of entities to external LOD authorities using LOD services and associated search APIs.

Beyond the *Emblematica Online* corpus, other digital library collections have been used to experiment with reconciling authority headings against unique local thesauri using external tools such as OpenRefine.[14] These efforts include developing unique URL-generating applications in various formats for name entities, ongoing maintenance of local controlled vocabularies, and metadata reconciliation practices.[15] This has yielded positive results such as high match rates and personal name tracings not found in LC authority files and are "the first steps toward a more integrated conceptualization of authority work."[16]

While efforts have been made to create local controlled vocabularies to provide standardized terms for individual digital libraries, this approach is most "advantageous when digital collections use shared controlled vocabularies" or when objects in digital collections are unique to local institutions.[17] For entities with an existing authority record that was established following the standards organizations such as LC, linking to existing sources of controlled vocabularies provides the additional advantage of matching a greater

percentage of terms.[18] In a 2019 project to prototype and test data models for the LOD environment, the University of Maryland Libraries enhanced the local corporate name authorities by reconciling with the LC Name Authority File (LCNAF) where possible, taking advantage of the existing external data.[19] Beyond individual library collections, collaborative work across libraries is also being deployed to reconcile field objects against existing LOD authorities to prepare library data at large for a transition from MARC format to Linked Data. In the most recent Program for Cooperative Cataloging's (PCC) work on changes in MARC encoding to accommodate LOD identifiers, MARC field objects are reconciled to RDF URLs from VIAF and LCNAF as part of the process.[20]

Some literature emphasizes the importance of metadata clean up before the reconciliation. Van Hooland et al. state, "Before asking the question of how to link metadata from different sources, we need to develop strategies to check their initial quality and possibly solve issues that might disturb the reconciliation process among different resources."[21] Southwick indicated in a 2015 study on transforming digital collections metadata into LOD that the implementation process would be more efficient if metadata clean up was done to the extent possible before reconciling with LOD sets.[22]

Other suggestions from the professional literature recommend that existing LOD authorities and services also present challenges, and may not always be sufficient substitutes for more traditional library authorities. This issue was also of interest to the authors as they conducted their case study. In a 2013 study to determine which controlled vocabularies were best suited for use in a scientific data repository, White quoted the findings from 2007 preliminary research that there was not a single vocabulary adequate to describe an interdisciplinary field such as evolutionary biology.[23] The same gap existed, and still does, for libraries in general. A 2016 study by Radio and Hanrath addresses the issue of inadequate subject representation as affecting a resource's ability to interact with the LOD environment.[24] They call for increased attention and participation to identify areas of under- or misrepresentation in Linked Data vocabularies.[25] Whereas the current body of literature is focused on examining workflows and procedures for metadata reconciliation to LOD, there remains a need for more research examining, assessing, and reporting on the efficacy of the reconciliation services and tools used (as measured by final match result).

Method

This case study used a hybrid methodology that consisted of quantitative analysis and qualitative comparison to accomplish the name and subject heading reconciliation process for the *Emblematica Online* collection data. For the quantitative analysis, the authors dissected the XML MODS/SPINE bibliographic records and identified name and subject heading strings as entities for reconciliation. They retrieved a subset of items from the corpus and examined the efficacy of the VIAF Auto Suggest API and LC Linked Data APIs in matching the name and subject heading entities respectively.

VIAF and LC authorities have been extended to provide LOD services and APIs that increase the usefulness of these authorities. VIAF and LC authorities were selected as the LOD resources to which to reconcile name and subject heading entities in this case study due to the applicability of their scope and the extensiveness of Linked Data services they provide. The aforementioned preexisting Python script was adapted to generate statistics on the name or subject authoritative links to which each entity was matched in one query. The VIAF Auto Suggest API provides a fast lookup for authority records in VIAF and returns JSON blocks of personal or corporate name records with the *viafid* included as a unique identifier. Based on the granularity of the queried name string, the query can return one result, multiple results, or none. For example, when a name string lacks birth and/or death dates, the query can return multiple JSON blocks of different name authority records because these name entities cannot be disambiguated. For the purpose of accuracy, the Python script counts a match when only one *viafid* was found in the returned RDF (JSON serialization).

> VIAF Auto Suggest API: http://www.viaf.org/viaf /AutoSuggest?query=[query string]

For subject heading entities, this study identified multiple LC controlled vocabularies as the LOD authorities for different types of subject headings in the original metadata. These authorities included LCNAF for name subject headings, LCSH for topical subject headings, Library of Congress Genre/Form Terms (LCGFT) for genre subject headings, and MARC Geographic Areas (GAC) for geographic subject headings. Multiple base URLs/APIs were therefore constructed accordingly to reconcile the subject heading entities:

- LCSH search API for topical subject headings: http://id.loc.gov/search/?q=scheme:http://id.loc.gov /authorities/subjects&q=aLabel: "[query string]"
- LCGFT search API for genre subject headings: http://id.loc.gov/search/?q=scheme:http://id.loc.gov /authorities/genreForms&q=aLabel: "[query string]"
- LCNAF search API for name subject headings: http://id.loc.gov/search/?q=scheme:http://id.loc.gov

/authorities/names&q=aLabel:"[query string]"

- GAC search API for geographic subject headings: http://id.loc.gov/search/?q=scheme:http://id.loc.gov/vocabulary/geographicAreas&q=aLabel:"[query string]"

As previously noted, the *aLabel* search constraint only returns a result that exactly matches the term searched.[26] Therefore, the match number for each subject heading entity using the LC Linked Data APIs will be either zero (not matched) or one (matched).

In addition to a quantitative analysis, the study included a qualitative comparative analysis based on an interview with Deren Kudeki, HathiTrust Research Center (HTRC) Developer at University of Illinois at Urbana-Champaign who had done parallel work with HathiTrust catalog records, to better understand the use of these APIs in another institutional context. Following the examination and analyses of match rates of name and subject heading entities using the search APIs, this study intended to suggest (for subsequent research and confirmation) implementation techniques and solutions that improve reconciliation match results.

Reconciliation and Enrichment Using APIs

For *Emblematica Online*, book and emblem catalog records are stored in XML MODS/SPINE format and are freely retrievable across the web. To implement LOD, the name and subject heading entities in the original XML metadata were enriched with VIAF and LC authoritative links, and the XML was transformed (using XSLT) and saved as RDF (JSON-LD serialization). As noted, this was made possible by the integration of VIAF Auto Suggest API and LC Linked Data APIs. To examine the efficacy of these APIs as metadata reconciliation tools, a quantitative analysis was conducted by retrieving a subset of XML files from each of the six major institutions that participate in *Emblematica Online*. Since the University of Illinois at Urbana-Champaign (UIUC) and the Herzog August Bibliothek (HAB) hold most of the XML files (together 63 percent of the corpus), the study retrieved more files from these two institutions' collections than the other four. For name entities specifically, fifty metadata files (emblem books) each from UIUC and HAB collections and ten metadata files each from the Glasgow University, Utrecht University, Duke University, and the Getty Research Institute collections were randomly retrieved. For the subject heading entities, fifty metadata files from the UIUC collection and ten metadata files each from the Glasgow University, Duke University, and the Getty Research Institute collections were randomly retrieved. LCSH is not present in the original HAB or Utrecht University metadata files because of the limits on consensus to use LCSH by all partners from different nations, so subject metadata from these two institution collections was not used for the subject heading analysis.

The Python script incorporates *matchCount* as a new variable to track the number of matched authoritative link(s) by the VIAF Auto Suggest API or LC Linked Data APIs for a certain entity, and writes the results to a CSV file. When a name or subject heading entity is queried, the script first uses an *if* statement to check whether a VIAF or LC authoritative link (*valueURI*) is already present for that entity in the original XML metadata (i.e., previously reconciled). If a *valueURI* is present, the algorithm will skip that entity and move to query the next. This helps avoid skewed results regarding the efficacy of APIs by excluding entities that were previously reconciled. One exception is the name entities in the HAB collection. The majority of the name entities in the HAB collection have previously reconciled *valueURIs* that points to the Deutsche National Bibliothek (DNB) authorities. Since DNB is not within the scope of this study, the algorithm ignores DNB *valueURIs* and queries the name entities in the HAB collection using the VIAF Auto Suggest API. As mentioned, the script counts a match when only one result was returned (*matchCount* = 1).

End Results of Entity Match Counts

Name Entities

Table 1 shows the number of unique name entities in the retrieved metadata files from each institution collection. Table 2 summarizes the number of queried name entities, number of unique match counts, and calculates the match rates.

One thing to note is that the number of name entities that were actually queried ("Number Entities Queried" in table 2) equals the Unique Name Entities (in table 1) less the number of name entities that already have a *valueURI* in the original metadata file. This step is necessary to avoid skewed results. For example, the script found 118 unique name entities in the UIUC sample, among which 22 already have a *valueURI*. The algorithm skipped those 22 and queried the remaining 96 name entities, on which the calculation of match rate is based. However, this does not apply to the name entities from the HAB collection, since the algorithm was intentionally designed to query HAB name entities from a non-DNB name authority—VIAF. Therefore, the number of queried HAB name entities (267) remains the same.

As shown in table 2, only one name entity was queried and matched for the Utrecht sample. The sample size is

Table 1. Number of unique name entities

Collection	Files Processed	Unique Name Entities
HAB	50	267
UIUC	50	118
Duke University	10	12
Getty Research Institute	10	25
Glasgow University	10	13
Utrecht University	10	13
TOTAL	**140**	**448**

Table 3. Number of subject heading entities

Collection	Files Processed	Unique Subject Heading Entities
HAB	0	0
UIUC	50	132
Duke University	10	13
Getty Research Institute	10	29
Glasgow University	10	13
Utrecht University	0	0
TOTAL	**80**	**187**

Table 2. Match rate of name entities using VIAF Auto Suggest API

Collection	Name Entities Queried	Match Count	Match Rate %
HAB	267	40	14.98
UIUC	96	60	62.50
Duke University	12	8	66.67
Getty Research Institute	18	9	50.00
Glasgow University	9	5	55.56
Utrecht University	1	1	100.00
TOTAL	**403**	**123**	**30.52**

Table 4. Match rate of subject heading entities using LC Linked Data APIs

Collection	Subject Heading Entities Queried	Match Count	Match Rate %
HAB	0	0	N/A
UIUC	129	54	41.86
Duke University	13	7	53.85
Getty Research Institute	29	14	48.28
Glasgow University	13	10	76.92
Utrecht University	0	0	N/A
TOTAL	**184**	**85**	**46.20**

too small for the 100 percent match rate to be statistically meaningful. The match rate for the HAB sample (14.98 percent) is noticeably lower than others. One possible reason that may contribute to this low match rate is how many of the name entities in the HAB collection are formatted. They are formatted as name acronyms (which tend not to return matches) instead of full names. Another reason for match failure likely is that the lack of birth and/or death dates in many of the name entities returns too many results. As mentioned, match counts greater than 1 are not considered a successful match. Reason for the lack of dates in HAB name strings is unclear, but could be due in part to differences in metadata formatting and cataloging practices in Germany versus the US. As aforementioned, because of the differences in cataloging and metadata practices across nations (the partners participating in *Emblematica Online* span four countries), there were limitations on the guidelines that could be established for the participating partners to follow when creating the original metadata. That being said, the adoption of LOD gives potential for improving the consistency and richness of metadata with global authoritative links that provide end-users with disambiguated and enriched information. In this case study, because of the nuances in the original metadata, the match rate of name entities in the HAB sample is not very representative of

the efficacy of the VIAF Auto Suggest API. Besides the highest match rate (100.00 percent for Utrecht) and the lowest (14.98 percent for HAB), the match rates of name entities using VIAF Auto Suggest API are between 50.00 percent-66.67 percent, with an average of 60.74 percent. With HAB included, the average match rate drops to 30.52 percent. This low HAB match rate suggests a need for further research as to why and to determine if a workaround is possible.

Subject Heading Entities

Table 3 presents the number of unique subject heading entities in the retrieved metadata files from each institution collection. Table 4 shows the number of queried subject heading entities, number of unique match counts, and calculates the match rates.

Similar to name entities and to avoid skewed results, the number of subject heading entities that were actually queried ("Subject Heading Entities Queried" in table 4) equals the Unique Subject Heading Entities in table 3 less the number of subject heading entities that already have a *valueURI* in the original metadata file. Since LCSH is not present in the original HAB and Utrecht University metadata files, no subject metadata from these two institution

collections was used for the subject heading analysis. Table 4 shows that the match rates of subject heading entities using LC Linked Data APIs are between 41.86 percent-76.92 percent, with an average of 46.20 percent.

HathiTrust Research Center (HTRC) LOD Project

As a reality check and to better appreciate the facets of this corpus that might influence reconciliation, the authors compared their results to those found for the HathiTrust Research Center (HTRC) LOD Project. The HTRC project works with metadata describing 17 million volumes across different institution's libraries to create BIBFRAME records from MARC records. The project reconciles contributor names to VIAF and subject headings to LC authorities. The same open source APIs and services are used—VIAF Auto Suggest API for name entities and LC Linked Data APIs for subject heading entities. Within the scope of LC controlled vocabularies, GAC is searched for geographic subject heading entities and LCSH for other subject headings. In 2019, name entities of over 17 million HTRC volumes had a match rate of 75.00 percent from VIAF, and subject heading entities of the same corpus had a match rate of 15.00 percent.[27] HTRC's match rate of 75.00 percent for its name entities is higher than the average match rate of name entities in this case study (60.74 percent not including HAB match rate, 30.52 percent including HAB match rate). This high match rate of the HTRC project was achieved by "using different ways to finesse the queries such as getting rid of the parentheses, and trying both a full date and just the start year in date."[28] The match rate for subject heading entities of the HTRC project (15.00 percent) is lower than the average match rate of subject heading entities in this case study (46.20 percent). Based on the interview and the authors' observations, they extrapolate some of the explanations for this difference:

- As a specialized collection, the subject headings in the *Emblematica Online* corpus are more uniform, such as *"Emblems," "Conduct of life," "Love in art,"* etc. that already have an established heading in the LC authorities. Subject headings in the HTRC corpus, in contrast, are much broader, with more than 17 million volumes on various subjects. It is possible that LC's Linked Data APIs respond better to specialized collections in reconciling subject heading entities, but more work is needed to prove that point.
- The HTRC project reconciled its general subject heading entities to LCSH and geographic subject heading entities to GAC. By contrast, *Emblematica Online* expanded to include LCGFT, LCNAF, and MARC Countries as part of the LOD authorities

used for the reconciliation in addition to LCSH and GAC. The use of multiple LOD authorities improved the match rate by matching genre and name subject heading entities to authoritative links that do not exist in LCSH or GAC.

Discussion

To transform digital library collection metadata into Linked Data, it is essential to implement a successful reconciliation that finds the best match to authoritative links for name and subject entities. Lessons learned from and the challenges during the reconciliation process of this case study are discussed below.

Prep Work before Reconciliation

It is important to minimize the metadata errors in the original metadata files. For example, one name entity in the HAB collection "*a. b. c. d. e. f. g. h. i. k. l. m. n. o. p. q. r. s. t. u. w. x. y. z.*" was erroneously recorded and returned no match result in VIAF. The correct form of the name in VIAF is "*A a b c d e f g h i k l m n o p q r r s s t u w x y z.*"[29] Several letters ("A," "r," and "s") were missing in the original name string. As a result, the authoritative link was not found by VIAF's Auto Suggest API. It is also important to ensure that the data is formatted correctly during processing, especially for non-English texts that involve diacritics. For example, one name entity "*Mabre Cramoisy, Sébastien*" was stored as "*Mabre Cramoisy, Sâebastien*" in the original XML metadata, which returned no match in VIAF because the Unicode character "*é*" was mistakenly transformed to "*âe*" during the data ingestion from the institution to *Emblematica Online*.

Data heterogeneity remains a challenge for metadata cleanup. It is hard to maintain consistency for heterogeneous digital collections when metadata is integrated from different sources or various data providers, as is the case for *Emblematica Online*. Van Hooland et al. pointed out that metadata quality and inconsistency will continue to remain a challenge for the reconciliation to LOD due to a lack of established methodologies or tools for metadata quality evaluation.[30] Specific to this case study, more consistent and standardized metadata would also have required more manual work on the legacy metadata and reaching consensus about matters of practice that have long varied across national boundaries. Metadata errors and incorrect ingested data in this case study were greatly minimized by the long-standing collaborations among the partners that led to the adoption of the SPINE schema, MODS usage guidelines, and high-level data content standards. Even so, as described above, enough variability in metadata

remained to create some challenges that interfere with the reconciliation process using the same API.

During Reconciliation

Discovering techniques to manipulate and format data strings is often needed to improve the match rate. During the reconciliation process of *Emblematica Online*, the authors experimented with two techniques to prepare metadata in a way that was proven to help find a unique match in VIAF:

- Changing angle brackets to square brackets. For example, no match was returned when using VIAF's Auto Suggest API for the name entity "*Sibylla Ursula <Braunschweig-Lüneburg, Herzogin>*" that was in the original metadata, but a match was found when the angle brackets were changed to square brackets and querying "*Sibylla Ursula [Braunschweig-Lüneburg, Herzogin]*".
- Removing punctuation at the end of a name string. For example, a unique result was returned for the name entity "*Mello, G. de*" but not for "*Mello, G. de.*".

However, it is worth noting that these formatting techniques vary by name and are difficult to anticipate in code. Depending on the LOD authorities and how the entity is formatted in that authority, one technique typically cannot apply to all entities across diverse collections (i.e., with metadata from diverse sources). For example, the name entity "*Josephus <Romanorum, Rex, I.>*" does not have a match in VIAF with either angle or square brackets. Similarly, "*Mauclerc, Antonius.*" returns a unique match result regardless of whether the period is present at the end of the string. The inconsistency of these formatting techniques presents challenges in preparing original or legacy metadata for reconciliation because there is no single solution to various formatting issues. As a result, it is up to the libraries and LOD practitioners to discover and implement what works best for their collection data.

LOD Resources as Authorities

It might be every LOD practitioner's dream that a single LOD authority contains all quality authority records that can be easily reconciled to by various entities. When White quoted the preliminary research conducted in 2007 that no single vocabulary was adequate for describing an interdisciplinary field, it was not clear that the same issue would be exemplified in today's ever-growing LOD implementation attempts.[31] For example, in LC Linked Data Service, geographic names are established in LCNAF, GAC, and MARC Countries, but not in LCSH.[32] This

means that to automate reconciliation of a geographic name entity used as a subject heading, LOD practitioners need to query other controlled vocabularies different from LCSH, such as LCNAF or GAC, to find a match to the authoritative link in LC Linked Data Service. By contrast, in traditional authority control practice, geographic names that can be assigned as geographic subject divisions can be easily searched manually by librarians in both "Name Authority Headings" and "Subject Authority Headings" using the LC authorities interface.[33] The ambiguity and inconsistency in how LOD resources connect to traditional library authorities like the LC authorities presents a challenge, and raises the question of whether LOD resources can be considered as encompassing the function and role of traditional library authorities.

Conclusion

This study describes the reconciliation of name and subject heading entities of *Emblematica Online* and examines the efficacy of the VIAF Auto Suggest API and LC Linked Data APIs in reconciling metadata to LOD authorities. Results from the quantitative analysis indicate that the average match rate of name entities using VIAF Auto Suggest API is 60.74 percent (without HAB match rate), and 30.52 percent (with HAB match rate). The average match rate of subject heading entities using LC Linked Data APIs is 46.20 percent. This study identifies solutions to improve match results of the metadata reconciliation in three aspects—data cleanup, formatting metadata strings, and paying attention to the ambiguity and inconsistency in how LOD resources connect to traditional library authorities.

The authors' case study adds to the growing body of work examining the application of LOD best practices to library special collections. The findings on the efficacy of VIAF Auto Suggest API and LC Linked Data APIs and the lessons learned through the course of this work can potentially be useful to personnel managing other digital libraries who are contemplating similar LOD reconciliation projects. Implementation tools and techniques in this study are easy to use and could provide opportunities for the larger digital library community to engage in incorporating LOD into the catalog.

However, the corpus used in this case study is limited to one specialized digital collection and only a small portion of the total corpus data was examined. A subsequent phase of research should extend the approach used here to the records of the entire corpus, refining the current approach to enhance the reconciliation match results. One possible direction for increased experimentation on this corpus would be to compare the scope and coverage of different LOD resources such as Wikidata, the Getty Art and

Architecture Thesaurus (AAT), the Bibliothèque nationale de France (BnF authority file), etc. Also, although the current approach yielded good reconciliation results for most institution collections in the *Emblematica Online* corpus, they did not work well for certain institutions. For example, the match result for name entities in the HAB collections using the VIAF Auto Suggest API was significantly lower than that of the other institution collections. The reasons for this need to be investigated further in a subsequent phase of work. This paper speculated the possible reasons based on observations, but it also shows the need to investigate the systematic disparity among different institution collections that would affect the final reconciliation results.

References and Notes

1. *The Princeton Encyclopedia of Poetry and Poetics*, 4th ed. (Princeton, NJ: Princeton University Press, 2012), s.v. "EMBLEM."

2. "Metadata Object Description Schema," Library of Congress, last modified February 5, 2020, http://www.loc.gov/standards/mods/; Stephen Rawles, "A Spine of Information Headings for Emblem-Related Electronic Resources," in *Digital Collections and the Management of Knowledge: Renaissance Emblem Literature as a Case Study for the Digitization of Rare Texts and Images*, ed. Mara R. Wade (Salzburg: DigiCULT Project, 2004), 19–28, https://www.digicult.info/pages/special.php.

3. Nuala Koetter, "Interoperability of Digital Emblematica Metadata Using the Open Archives Initiative Metadata Harvesting Protocol and Other Schemas," in *Digital Collections and the Management of Knowledge: Renaissance Emblem Literature as a Case Study for the Digitization of Rare Texts and Images*, ed. Mara R. Wade (Salzburg: DigiCULT Project, 2004), 79–87, https://www.digicult.info/pages/special.php; Thomas Stäcker, "Transporting Emblem Metadata with OAI," ed. Mara R. Wade (Salzburg: DigiCULT Project, 2004), 89–95, https://www.digicult.info/pages/special.php.

4. "EmblematicaOnline," accessed May 28, 2021, https://github.com/cindyttian/EmblematicaOnline/blob/main/end2end-rev5Dec2019.py.

5. *Linked Data Glossary*, s.v. "Entity," by W3C Working Group Note, last modified June 27, 2013, https://www.w3.org/TR/ld-glossary.

6. "VIAF—Authority Cluster Auto Suggest," OCLC Developer Network, accessed December 2, 2020, https://platform.worldcat.org/api-explorer/apis/VIAF/AuthorityCluster/AutoSuggest; "Technical Center," Library of Congress Linked Data Service, accessed December 2, 2020, https://id.loc.gov/techcenter/searching.html.

7. "VIAF—Authority Cluster Auto Suggest."

8. Katrina Fenlon et al., "Exploring Linked Data Benefits for Digital Library Users," *Proceedings of the Association for Information Science & Technology* 55, no. 1 (2018): 799–800, https://doi.org/10.1002/pra2.2018.14505501122.

9. Tim Berners-Lee, "Linked Data," July 27, 2006, https://www.w3.org/DesignIssues/LinkedData.html.

10. Jeremy Myntti and Anna Neatrour, "Use Existing Data First: Reconcile Metadata before Creating New Controlled Vocabularies," *Journal of Library Metadata* 15, no. 3–4 (2015): 205, https://doi.org/10.1080/19386389.2015.1099989.

11. Timothy W. Cole et al., "Library MARC Records into Linked Open Data: Challenges and Opportunities," *Journal of Library Metadata* 13, no. 2–3 (2013): 163–96, https://doi.org/10.1080/19386389.2013.826074.

12. Timothy W. Cole et al., "Using Linked Open Data to Enhance the Discoverability, Functionality and Impact of Emblematica Online," *Library Hi Tech* 35, no. 1 (2017): 159–78, https://doi.org/10.1108/LHT-11-2016-0126.

13. Cole et al., "Using Linked Open Data."

14. Scott Carlson and Amber Seely, "Using OpenRefine's Reconciliation to Validate Local Authority Headings," *Cataloging & Classification Quarterly* 55, no. 1 (2017): 1–11, https://doi.org/10.1080/01639374.2016.1245693; Silvia B. Southwick, "A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies," *Journal of Library Metadata* 15, no. 1 (2015): 1–35, https://doi.org/10.1080/19386389.2015.1007009.

15. "django-name," accessed October 19, 2020, https://github.com/unt-libraries/django-name; Jeannette Ho, "Name Disambiguation for Digital Collections: Planning a Linked Data App for Authority Control at Texas A&M University Libraries" (presentation at the LD4 Conference on Linked Data in Libraries, Boston, MA, May 11, 2019).

16. Carlson and Seely, "Using OpenRefine's Reconciliation," 10.

17. Southwick, "A Guide for Transforming Digital Collections Metadata," 20.

18. Myntti and Neatrour, "Use Existing Data First," 191–207.

19. Bria Parker and Adam Gray, "Rethinking the University of Maryland Authority File for the LOD Environment," *Journal of Library Metadata* 19, no. 1–2 (2019): 69–81, https://doi.org/10.1080/19386389.2019.1589699.

20. Jackie Shieh, "PCC's Work on URIs in MARC," *Cataloging & Classification Quarterly* 58, no. 3–4 (2020): 418–27, https://doi.org/10.1080/01639374.2019.1705951.

21. Seth van Hooland et al., "Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections," *Journal of the American Society for*

Information Science and Technology 64, no. 3 (2013): 469, https://doi.org/10.1002/asi.22763.

22. Southwick, "A Guide for Transforming Digital Collections Metadata," 19.

23. Hollie White, "Examining Scientific Vocabulary: Mapping Controlled Vocabularies with Free Text Keywords," *Cataloging & Classification Quarterly* 51, no. 6 (2013): 655–74, https://doi.org/10.1080/01639374.2013.777004.

24. Erik Radio and Scott Hanrath, "Measuring the Impact and Effectiveness of Transitioning to a Linked Data Vocabulary," *Journal of Library Metadata* 16, no. 2 (2016): 80–94, https://doi.org/10.1080/19386389.2016.1215734.

25. Radio and Hanrath, "Measuring the Impact and Effectiveness of Transitioning to a Linked Data Vocabulary," 92.

26. "Technical Center," Library of Congress Linked Data Service, accessed December 2, 2020, https://id.loc.gov/techcenter/searching.html.

27. Deren Kudeki (HathiTrust Research Center Developer at University of Illinois at Urbana-Champaign), interview with author, November 15, 2019.

28. Kudeki interview with author, November 15, 2019.

29. "A a b c d e f g h i k l m n o p q r r s s t u w x y z," VIAF, accessed October 11, 2020, http://viaf.org/viaf/33061207.

30. Van Hooland et al., "Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections," 469.

31. White, "Examining Scientific Vocabulary," 658.

32. "Library of Congress Names," Library of Congress Linked Data Service, accessed June 8, 2021, https://id.loc.gov/authorities/names.html; "MARC List for Geographic Areas," Library of Congress Linked Data Service, accessed June 8, 2021, https://id.loc.gov/vocabulary/geographicAreas.html; "Library of Congress Subject Headings," Library of Congress Linked Data Service, accessed June 8, 2021, https://id.loc.gov/authorities/subjects.html.

33. Library of Congress Authorities, accessed June 8, 2021, https://authorities.loc.gov/.

Dispersed Collections in Exile

Thai Collections in Libraries outside of Thailand

Hollie White and Songphan Choemprayong

A previous research conducted by White and Choemprayong in 2019 indicates that libraries in Thailand have encountered multidimensional challenges in contributing to global bibliographic and resource sharing networks. As libraries outside Thailand may have a strong influence on representation in the global bibliographic universe, this paper investigates how libraries outside of Thailand manage Thai collections, particularly acquisitions and cataloging. Adopting Merton's Insiders/ Outsiders doctrine as a contextual framework, this study applies an exploratory mixed-method approach using observation and interview methods.

S haring through library lending and cooperative cataloging systems has become a normative approach to improve the efficiency of collection management and to enhance the accessibility of collections. To support collection sharing on a global scale, library communities have standardized collection management and description practices for decades. However, these standards have been criticized in terms of biases, particularly the challenges of acquiring and cataloging collections representing marginalized populations or topics.[1]

Standard or central Thai is Thailand's official language, and "is the language used at all levels of education and has been so almost exclusively for about one hundred years."[2] While Thai people have migrated elsewhere around the world (e.g., Thai diaspora), the collections written in the Thai language are mainly produced and distributed from within Thailand itself. Lee observed that there was a shortage of academic content written in the Thai language.[3] Even within Thailand-based institutions, a number of academic libraries have a larger collection in foreign languages, especially English, than in Thai. Previous studies have found that Thai libraries face numerous obstacles in participating and contributing to the global bibliographic and resource sharing networks, including inefficiency from maintaining multiple versions of bibliographic records, additional efforts for transliteration, budget limitations, and challenges integrating content into OCLC.[4]

Nonetheless, the collection of Thai materials in institutions outside of Thailand has occurred since the seventeenth century.[5] Thai manuscripts, publications, and artifacts can be found in libraries, museums, and collectors' private collections around the world. Additionally, Thai studies, as a multidisciplinary field of study, has existed since 1904.[6] Research interests in Thai studies have been extended to scholars in many countries within and outside of Thailand. Libraries outside of Thailand collect Thai language materials for two reasons: to serve the needs of the Thai speaking diaspora and to support Thai Studies

Hollie White (hollie.white@curtin.edu.au) is a Senior Lecturer in the School of Media, Creative Arts and Social Inquiry at Curtin University in Perth, Australia. **Songphan Choemprayong** (songphan.c@chula .ac.th) is an assistant professor at the Department of Library Science, and a principal researcher at the Arc of Memory Research Unit, Faculty of Arts, Chulalongkorn University in Bangkok, Thailand.

researchers. Changes to library budgets and collections in the twenty-first century resulted in less emphasis on print material and more focus on nationalistic approaches to collection building and maintenance.[7] This study investigates the extent to which Thai materials have been collected and described in libraries outside Thailand.

Literature Review

Collection Management of Thai Materials

In general, Thai collections outside of Thailand can be divided into two major categories: manuscripts and contemporary publications. Collecting of Thai manuscripts in Europe started in the early seventeenth century.[8] Manuscripts and documents were originally taken outside of Thailand by traders, travelers, and soldiers and later by scholars, missionaries, administrators, and commercial agents.[9] According to McDaniel, "when travelers from Europe, North America, and Japan, among other places, started exploring Southeast Asia, they often brought manuscripts back to their own museums and homes."[10] Manuscript collections of Thai materials outside of Thailand usually focus on content from the nineteenth century with major collections located in Germany, Great Britain, Ireland, Italy, Japan, and the United States.[11] McDaniel argues that most Thai manuscripts were brought to other countries as part of diplomatic exchanges or royal gifts, and despite ending up in university special collections, these items were not purchased as part of a larger collection development plan.[12]

A few sources agree that it was not until after World War II that collectors and librarians outside of Thailand started to actively seek out Thai materials.[13] This could be for two reasons: (1) "The decade from 1947 was critical in both reinforcing the need to develop resources for the study of Asia and in implementing collecting policies. UNESCO was particularly active in promoting cultural exchanges, including the development of exchanges between libraries," and (2) American-based Asian studies programs in the 1950s were well resourced and funded.[14] The Southeast Asian book trade from the late 1940s until the end of the 1970s was difficult because it was "unorganised, undercapitalised, and relatively inefficient."[15] Other changes from the post-World War II collecting era show a shift of knowing little about the person or people who brought the collection originally from outside of Thailand to having detailed information about collections and donors.[16]

As previously noted, donations from private collections or gifts provide one source of materials.[17] Roughly 650 Thai manuscripts are held in public collections in the United States.[18] British institutions house around 500 Thai manuscripts.[19] Thailand and Japan had a relationship to share

Buddhist texts in the seventeenth century and then again in the early twentieth century, and there are a number of Thai items at Japanese temples or universities.[20] Thai materials in Germany are sparse and scattered, with many Thai items being unidentified and uncatalogued.[21] There is no single resource that points to all the major Thai collections (manuscript or otherwise) outside of Thailand.[22] Information about contemporary Thai collections is less abundant than that of Thai manuscript collections. Many sources of information about contemporary Thai collections are available on institutional websites and in LibGuides. Moreover, national libraries and consortia can play an important role in the collection of contemporary materials by creating connections and programs to support the larger library collecting community.

National Libraries Role in Collecting Thai Materials

National libraries, such as the British Library, the Library of Congress (LC), and National Library of Australia, can play an important part in collecting materials from other countries. Based on a 2015 International Federation of Library Association (IFLA) National Libraries Section Study, the role of some national libraries in collecting "foreign" materials states,

> Some national libraries are tasked with collecting foreign material related to the country in question. In most developed countries the national library acquires and maintains large collections of foreign literature in most fields of research to provide encyclopedic reference services to the nation's researchers or to supplement research collections in other libraries. National library acquisitions mandate is influenced by national priorities in research, the availability of other collections in the framework of national provision, and the availability of material through alternative sources, for example, by connecting to content via online services.[23]

Based on the idea of the role of national libraries, a brief discussion follows the history of collecting Thai materials by three national libraries: the British Library, LC, and the National Library of Australia. By no means are these the only national libraries that collect Thai materials, as a number of other national libraries collect either manuscripts or contemporary Thai content. However, these western countries have a long term historical and political relationship with Thailand and strong visibility in terms of interest in collecting Thai materials.

The British Library's Thai collection is "the most

extensive and important collection" in the United Kingdom, and "one of the most significant in Europe."[24] Because long-established international bibliographic exchange programs, "in the early twentieth century, most publications by the Thai National Library (formerly Vajirañāṇa Library) were deposited at the British Museum Library."[25] Over time, collections in other British institutions were merged into the British Library's current Thai collection. For example, "although the first fine illustrated Thai manuscript came into the India Office Library in 1825 (MS Pali 207), and some other Thai manuscripts were transferred from the British Museum Library to the British Library in 1973, the Thai manuscript collection was mainly built up from 1973 onwards."[26] Most of the items in the Thai collection were gifts from Christian missionaries.[27] According to the British Library's Thailand, Laos and Cambodia collection website, information about Thai printed books published prior to 2000 are not cataloged or available online, and are only available via a physical card catalog found in the Asian and African Studies Reading Room.[28] Since 2008, the British Library has digitized many of their unique Thai pieces.[29]

LC has about 55,000 Thai language items consisting of both manuscript and contemporary collections. Additionally, LC has managed an overseas office in Jakarta, Indonesia since 1963, which coordinates the acquisition of materials from Brunei, Cambodia, Indonesia, Laos, Malaysia, Philippines, Singapore, Thailand, and Vietnam.[30] The overseas office also coordinates LC's Cooperative Acquisitions Program for Southeast Asia (CAPSEA) program and materials exchange programs. One of the Jakarta suboffices is in Bangkok, Thailand at the US Embassy. This sub-office acquires around 750 serials and 1,500 monographs annually.[31]

For many years, the National Library of Australia (NLA) "house[d] the largest and most actively developing research resource on Asia in Australia."[32] This includes a large collection of post-World War II Thai language-based materials. Thai collecting at the NLA started in the 1950s.[33] By the 21st century, NLA's Thai Collection housed more than 30,000 monographs and 100 current journal and popular magazine titles.[34] The NLA decided in late 2019/early 2020 that it would stop collecting in many Asian languages, including Thai materials due to shifting national priorities that impact budget.[35]

Consortia Roles in Collecting Thai Materials

In 1978, the National Libraries and Documentation Center of the Southeast Asia Consortium was founded to initiate the interlibrary exchange program among national libraries and documentation centers in Singapore, the Philippines, Malaysia, Indonesia, and Thailand.[36] However, there is no publicly available evidence describing how the network operates. In addition to collaboration among national libraries, there are two prominent international consortia that discuss and share information about collecting Thai materials and other Southeast Asian resources: the Southeast Asia Library Group (SEALG) in Europe, and the Committee on Research Materials on Southeast Asia (CORMOSEA) in North America. The CORMOSEA website emphasizes the important role that LC's Cooperative Acquisitions Program for Southeast Asia (CAPSEA) plays in assisting member institutions to build their collections.[37] Interestingly, CORMOSEA is a group of libraries that organizes under the Thai studies community, rather than a library-based association.

Thai Collections and Euro-American Bias

Libraries outside of Thailand play a role in creating access to information about Thai language and culture. Those libraries outside Thailand are uniquely positioned to collect materials and ephemera that may otherwise not be collected within Thailand itself. For example, this may include print publications or online content related to "shifts in politics and ideas," social movements, coups, or content critical of the monarchy that would not normally be collected and maintained in institutions within Thailand.[38] In another respect, Thai language collections outside Thailand represent a mainly Western lens or filtered view of Thailand and its aboutness.

Information about these collections emphasize a Euro-American centric value system that privileges white, Western experiences as adding validity to Thai materials, thus impacting how materials are acquired, described, and maintained. For example, "the British Library's Southeast Asian collections, were strongly based on and reflected British imperial involvement in the region," and reflects the personal interests of long-time curator Henry Ginsburg.[39] Furthermore, when discussing Thai collections outside Thailand, the donor, instead of the materials, are a huge focus of the collection information. Examples are apparent in presentations like Ashmun, and papers such as those by Walker and Kerekes and McDaniel.[40] These works emphasize how the donor experienced Thai culture, but not the content or materials that reflects the culture itself. These acts center on westernness and whiteness and not the Thai experiences, viewpoints and culture seen in the materials themselves.

Euro-American Bias in Description, Access, and Discovery of Thai Materials

The bias of knowledge organization systems, cataloging standards, and controlled vocabularies is widely known within the library community.[41] Many cataloging and metadata researchers have examined how the standards and

systems created to aid description and access in libraries are socially representative of the time and context in which they were created, and therefore not easily used outside those contexts. The problematic nature of the standards and systems used for description can impact users' ability to find and access the material they need.

Transliteration, also known as Romanization, is the conversion of a non-Latin script to a phonetic-based Latin script to assist in retrieval. Romanization continues to be a common practice in library discovery systems. Numerous attempts have been made to Romanize Thai script. The first attempts originated from French visitors trying to record the Thai language on paper based on what they heard using French linguistic patterns.[42] In the late nineteenth and early twentieth centuries, more structured attempts by Asian Studies scholars "distort[ed] the spelling of native Thai words," causing criticism from King Vajiravudh (Rama VI).[43] Thai attempts to create Romanization systems resulted in a practical system created in 1931, and the Royal Institute of Thailand published a Romanization guide in 1968.[44] Currently Thammasat University's Plangsarn transliteration system is widely used by libraries in Thailand.[45]

The first LC/American Library Association (LC/ALA) Romanization system was established in 1958 and has since been continuously updated.[46] According to LC, "Romanization is primarily for LC staff and staff at other libraries without language expertise working in: 1. Circulation 2. Acquisitions 3. Serials check-in, 4. Shelf listing, 5. Shelving 6. Reference. Romanization is also for systems that cannot use non-Latin forms, have support for only some scripts, or require Romanized fields for indexing and sorting purposes."[47] The most recent LC Thai Romanization table was published in 2011. LC's current documentation states that Romanization still "merits continued application and support," despite also acknowledging that "users who can read those languages and scripts (i.e., languages in their original scripts) appreciate our [LC] providing bibliographic information in those scripts."[48]

In February 2020, the Cornell University Library Usability Working Group conducted a study about patron use of the catalog to search for Thai language materials.[49] Cornell University Library is known for both its manuscript and contemporary Thai collections.[50] The study results indicated that "participants said that using Thai is important when searching but think that Romanization is useful to have."[51] Additionally, while "most participants used a standard Romanization scheme," the users did not use the LC standard that was implemented in the catalog, Instead, they used a different scheme, which was unidentified in the study results.[52]

According to Weinberg, the Romanization process in an era of advanced library systems is "a philosophy of Anglo-centricity vs. multiculturalism."[53] She argues that "fully Romanized bibliographic records coincide with *Anglo-centricity* in the United States—the view that everyone throughout the world should learn to speak the English language and master the alphabet in which it is written."[54] This privileging of Anglocentricity and whiteness has impacted how Thailand-based catalogers and librarians interact with international standards and systems.

As of December 2020, there are twenty-one Thailand-based OCLC member libraries and almost 1 million WorldCat.org records representing cataloged Thai language resources. The number of resources in these libraries outnumber the number of records shared within the WorldCat system. Although the purpose of the OCLC program is mainly to promote resource sharing among libraries worldwide, there seems to be barriers for Thai libraries to join and successfully contribute to the network.[55] White and Choemprayong identified that various standards and configurations of different local and international consortiums have become a source of burnout for Thailand-based librarians who view creating and editing bibliographic records in network platforms as an unnecessary process, consuming additional time and resources in the collection management and cataloging workflow.[56]

The contributions of Thai libraries in providing access to their holdings in any network system can be difficult to observe and estimate for several reasons. For example, while Thai catalogers appreciate the availability of cataloging copy, they are reluctant to update and share their own records in the network system.[57] Instead, Thai catalogers tend to edit and restrict records to their local catalogs and not share revisions. If updating is required, Thai catalogers typically create new records, rather than update the existing ones. Additionally, the Romanization of cataloging records is perceived as a great burden as it increases the time to process an individual record. The barriers to entry to the network systems among Thai libraries raises an essential concern regarding access to current and future Thai materials. With a low number of participating Thai libraries in global networking infrastructure, the majority of Thai materials represented in global networks are from institutions located outside Thailand. The purpose of the study presented here is to examine Thai language collections outside Thailand to understand where these collections are located and how they are being maintained and described.

Method

This study used a library and information science approach, as opposed to a Thai studies approach, for examining Thai language collections in libraries outside Thailand by focusing on library-based sources and perspectives related to collection maintenance. An exploratory mixed method

approach to examine Thai collections outside Thailand was used and expanded upon the Bangkok-focused research presented by White and Choemprayong, which focused on the cataloging practices of Thai librarians across library sectors.[58] An exploratory sequential approach was used to observe an issue/problem that are not clearly defined.[59] In this study, a quantitative content analysis of libraries websites, catalogs, and LibGuides was first conducted to identify characteristics of Thai materials outside Thailand and their collectors. A qualitative semi-structured interview was carried out subsequently to obtain rich information, including opinions and perspectives regarding the current practices and issues regarding managing these collections. This study, unlike the previous one in 2019, focuses more on collections while using catalog data and the global cataloging community to find and understand the ecosystem of Thai language collections outside of Thailand.

Conceptual Framework

This study focused on the practices and perspectives regarding collection management of resources that circulated within and transferred across libraries within and outside Thailand. Merton's social epistemological conceptualization of the Insiders/Outsider doctrine provides a conceptual lens to understand the worldviews of library collection management inside and outside Thailand.[60] Notably, Merton's conceptualization was the foundation of Chatman's notion of small world, a renowned information behavior framework.[61] In Merton's Insiders/Outsiders doctrine, members in a group with a "privilege access" to certain knowledge or information can be considered as the Insiders. In contrast, the Outsiders expend more effort, consume more resources, and encounter a higher risk to access to the same knowledge and information.[62] The segregation is based on the innate status (i.e., geographical location for this study), rather than merit. From an Insider's perspective, the Outsiders neither have a direct experience nor continuously engage in socialization activities. Thus, the Outsiders lack an "intuitive sensitivity," "shared realities," and "fine-grained meanings of behavior, feelings, and values."[63] The Outsiders have substantial distinguished aims and goals concerning issues relevant to their own values and interests. This doctrine also concerns the conditions of interactions between Insiders and Outsiders, particularly in the intellectual and ideological domains. For instance, in conflicts, holding predominating arguments (i.e., Insiders) that are accepted by ones who hold the opposite view (i.e., Outsiders) can be considered as a triumph of Insiders. However, the Outsiders who seek an acceptance by Insiders may portray a stronger passionate action and enthusiasm than the Insiders would do. While Merton argues that the social structure of Insiders and Outsiders in realities are dynamic, complex, and fluid, the

application of the doctrine in an ascribe dimension (herein inside and outside Thailand) would provide an in-depth interpretation framework on how libraries outside Thailand manage Thai collections.[64] Earlier studies in library and information science adopted Merton's doctrine to explain how information users seek and share information within and outside their social world.[65] In this study, Merton's notion was applied to libraries outside Thailand. In addition to the geographical location of the libraries, the social structure of interest includes the cultural identity of the library staff (i.e., Thai versus non-Thai) involved in Thai material acquisition and maintenance.

Research Question

This study's research question is "what are the current practices for Thai collecting and description in libraries outside of Thailand?" By answering this question, the researchers aimed to discover some libraries that currently hold and collect current Thai materials, how Thai collections are obtained and described, and to gain an understanding of the staff who work with these collections.

Finding Thai Materials Outside Thailand

Since there is no published exhaustive list of libraries that currently hold Thai materials worldwide, potential participants who worked with Thai collections in libraries outside Thailand were identified using purposive and snowball sampling techniques. The researchers consulted library websites, catalogs, and LibGuides as a main source to identify potential libraries with Thai collections. Furthermore, library networks specializing in Southeast Asia collections mostly in the United States and United Kingdom were also used to identify potential participant libraries. A general email invite was circulated on library cataloging discussion lists in Europe and North Africa/the Middle East. The National Library of Thailand Exchange List was also consulted to find potential participants.

Openly available online information on Thai collections was gathered from potential interview participants, including reputation about Thai specialization, number of volumes in the Thai language, and currency of the collection. Specific collection information was placed into three categories: special collections (such as historical collections and cremation volumes), manuscripts (such as palm leaf manuscripts), or current collections (actively collecting items published from 1950 onward). It is important to identify libraries that were actively and consciously collecting Thai materials because they would likely have policies and procedures specifically for Thai collections. Thus, if a collection was identified as having the following:

- a strong reputation (based on review of literature),
- evidence of Thai specialization: a collection of 1,000 or more current (1950 to present) titles in Thai language, and
- actively cataloging materials themselves,

an interview request was sent to a collection or cataloging library professional at that institution.

Data Collection

Data collection occurred between November 2019 through August 2020. First, openly available online information found on library websites, catalogs, and LibGuides (as noted in the sampling section previously) was used to answer basic questions related to characteristics of materials and the collecting libraries. Content collected for this section included institution name, potential contact, date contacted, notes about the Thai collection, region where the institution is located, plus the criteria listed in the sampling section above.

Website, catalog, or LibGuide data was collected from fifty-one libraries in North America, Asia, Europe, and Oceania. Libraries elsewhere were investigated, but Thai collections could not be identified in China, Russia, North Africa, the Middle East and South America. Attempts to contact librarians in these regions were not successful. After the initial online data was collected, all fifty-one libraries were contacted for collection clarification. A total of thirty-two libraries responded to initial inquiries related to the interview qualifications: reputation, Thai specialization, and actively managing the collection.

Nineteen libraries were excluded from the study. Reasons for exclusion include (1) the participant did not respond to initial contact, (2) the library no longer held Thai materials (i.e., the collections were transferred to other institutions), (3) the library lacked dedicated staff responsible for managing the Thai collections, (4) the library used outsourcing to process Thai collections, (5) the library does not acquire new Thai materials, and (6) the library holds only special small collections (e.g., rare books, personal collections, and language materials).

Thirteen libraries were invited to participate in interviews, with nine agreeing. Table 1 shows the number of participating libraries categorized by global region, both for online data collection and interview participants. Six libraries participated in the interviews via email. Two interviews were conducted in-person. One interview was conducted using video chat. Various types of libraries participated in the interview portion of the study, including three national libraries, four academic or research libraries, and two special libraries.

Semi-structured interviews were used to elaborate on practices and opinions about Thai material acquisition and maintenance. As approved by Curtin University's

Table 1. Number of Participating Libraries by Region

Region	Number of Library Participants (Online Data)	Number of Library Participants (Interview)
Asia	9	3
Europe	15	2
North America	21	2
Oceania	6	2
Total	**51**	**9**

Ethics Committee, the interviews were conducted either in person, via video call, or by email. Since it was impossible to visit all participants in-person, especially during the COVID-19 pandemic, the participants chose their preferred method of communication. All interview content is available in English or Thai. The interview questions used in all interview modes in this study are included in appendix A. Questions addressed cataloging related activities and based on responses, follow up questions were asked related to collections and acquisitions. Additionally, participants were allowed to provide other comments or observations regarding practices in Thai material acquisition and maintenance. For in-person interviews, audio recordings were used to capture the interview content and the researcher's notes. The interviews recorded using video calls were recorded. During the interviews, the researchers created memo notes that were used during the data analysis for data triangulation.

Data Analysis

Two researchers conducted a deductive thematic analysis of websites and interview data based on the RITA method, which involves rapid identification of themes as opposed to precise transcription.[66] The original theme was developed based on the research question and constructs of interest from relevant literature. Data analysis was iterative and conducted from August 2020 through December 2020. Emerging codes were added into the original coding scheme. The coding process was done manually. All codes were categorized into relevant themes. The resulting themes included collection characteristics, staffing, acquisition practices, and cataloging practices.

Data Reporting and Anonymity

This research is supported by Curtin University Ethics HRE2019-0684, which was approved in October 2019 and renewed in October 2020. Institution names will be included when data collection related to research question part A,

as it collocates information that is openly available online, is reported. The names of institutions and the individuals who were contacted will not be included in respect to the anonymity process of the data collection. All interview participants provided informed consent prior to the participation.

Results

Study results begin with providing collection characteristics in libraries outside Thailand, while still maintaining anonymity. This section discusses study findings based on three areas: staffing, acquisition practices, and cataloging practices.

Characteristics of Materials in Libraries Outside Thailand

Of the fifty-one libraries researched, twenty-six had some type of Thai collection, either in the past or present. Eleven participants had larger Asian region collections, while eight focused on Southeast Asia generally. Three libraries only collected in East Asia, and the last three had collections that contained random Asian materials or could not be clearly identified into the other Asian region classifications.

As twenty-six of the contacted libraries indicated having some type of Thai collections, many libraries had stopped collecting Thai materials for a variety of reasons including: a refocus of collection development policy, changing budgets, the institution no longer taught Thai or Thai studies, no longer having institution-based researchers who study Thai topics, or no longer having staff who could read Thai.

Types of Materials

Thai materials collected outside Thailand are generally diverse, but unsurprisingly similar to other library collections. There are a wide range of collection types in these participating libraries. Table 2 includes a list of resource types of Thai materials in libraries outside Thailand based on the study findings.

Staffing

One of the study's main findings focused on the variety of staffing issues in managing and describing Thai materials. For those libraries with in-house Thai language cataloging, the catalogers were typically born in Thailand and now live in the other country as Thai diaspora. Most Thai staff in libraries outside Thailand who participated in the interview do not hold a library and information science degree. Those staff normally have in-house, on the job training related to library practices. One European-based participant's library

has a branch office/library based in Bangkok. The staff located in Thailand were responsible for cataloging Thai materials using the centralized cataloging system.

Based on the interview responses, some libraries lack Thai staff in the collection development or cataloging team, and instead have non-ethnically Thai professional librarians working with Thai materials. A few librarians received additional training on Thai language in Thailand prior to working with Thai materials. Nevertheless, in some libraries, the local staff obtained the knowledge about working with Thai materials from the part-time Thai staff or previously full-time staff. These staff reportedly could not read Thai. Instead, they relied heavily on the sources of cataloging records, which will be explored in detail in the cataloging subsection. Yet a number of libraries reported not having someone on staff who can catalog Thai materials. In those cases, two approaches are used: outsourcing to vendors and contractors or leaving materials uncataloged or backlogged. One Oceania region participant uses their former retired librarian to catalog Thai materials on a part-time basis.

Acquisition Practices

Study findings show that Thai collections are acquired in various ways, depending on institutional budgets, and collecting priorities. Asian library participants indicated that Thai materials are available through local markets or vendors. However, most reported that there are only a small number of local distributors for Thai content, or even other Southeast Asian materials. Local sources were not listed as a major supplier for participant libraries outside of Asia.

Direct purchasing or exchanges from sources in Thailand was the most popular method used by study participants. Some of these purchases happened when librarians traveled to Thailand and purchased directly from stores during their trips. One North American participating

Table 2. Types of Thai Materials held in Collections outside of Thailand

Monograph	**Archival Materials**
• Textbooks	• Microfilms
• Research publications including theses and dissertations	• Palm leaf manuscripts
• Cremation volumes	• Rare books (as early as 1840 and colonial period)
• Donated personal collections	• Maps
• Buddhist text and tipitaka collections	
• Language learning materials	
Periodicals	**Recreational Materials**
• Journals	• Literature and fiction
• Magazines	• Films
• Newspapers	• Videos

library used a former employee who now lives in Thailand to purchase from stores in Thailand. Other participating libraries used agents or vendors based in Thailand. Library exchange programs with the National Library of Thailand was another source of materials. Collections that rely on the interests and connections of one particular researcher was observed in two cases (one in North America and one in Oceania). In one case, the entire collection growth was based on annual gifts from one particular professor with no other funding provided to grow the Thai collection.

Cataloging Practices

Three areas of cataloging practices are explored in this subsection: sources of cataloging records, describing collections, and assigning controlled vocabularies and classification.

Source of Cataloging Records

Most libraries relied on LC and OCLC to obtain bibliographic records of Thai materials. However, a few participants reported that bibliographic records for Thai materials, and other Southeast Asian collections, are sparse and unauthoritative in OCLC. In some Asian libraries, where Thai materials are strongly represented, national cataloging networks are used to obtain bibliographic records. Other libraries that were cited as sources of bibliographic records include the National Library of Australia, Thammasat University Library for Romanized data, and the National Library of Thailand's website for Thai language-based records.

Describing Collections

When libraries' materials were cataloged in-house, Resource Description and Access (RDA) was the commonly used standard among the participating libraries. Libraries may only apply RDA to new materials (published in 2013 or later), while earlier materials (published before 2013) were cataloged following the Anglo-American Cataloging Rules 2nd Revision (AACR2) format. A number of participating libraries use national-level cataloging rules to describe Thai periodicals. The application of certain MARC encoding schema is also dependent on local policies and standards (e.g., UNIMARC vs. MARC21). Romanization and when to apply it was a main source of discussion in relation to challenges with description.

Romanization Issues Related to Thai Materials

Finding records for Thai materials in OCLC was identified as a major challenge for interview participants. An issue that may contribute to this challenge is how titles are Romanized. As cited in the literature review and findings sections, Romanization of Thai in general is a difficult task due to numerous reasons because

> describing the title of books in accordance with the LC Romanization method is the one of the most challenging works. Thai Romanization table, especially Thai word division has many exceptions (16 pages in total), therefore, describing Romanization properly is almost impossible unless we understand the Thai Romanization table correctly and are familiar with Thai pronunciation. "Plaengsan" (http://164.115.23.167/plangsarn/) provided by Thammasat University Library is a useful Romanization tool, but sometimes it has mistakes, so we should modify them in usual.[67]

Interview results revealed four variations in Thai title descriptions and deciding whether to Romanize. Those were (1) Romanized title, (2) original Thai spelling, (3) original Thai spelling with tokenized terms, and (4) translated title. For Romanization of Thai titles, two systems are used: the ALA-LC Romanization table and Thammasat University's Romanization table. While LC records disregard the original Thai spelling, some libraries reportedly use both variations in describing the title. Some choose to provide only the Romanized title in bibliographic records. Other libraries provide descriptions and indexes in the native language in combination with English. One participant explained that the Romanization of Thai for materials with Buddhism-related content is the most difficult task since the title normally contains both Thai and Pali. The cataloger must decide whether to use the Romanization of Thai or of Pali, with little choice to use both.

Additionally, the structure of the Thai language presents a challenge for retrieval. In Thai script, a sentence or phrase is normally written without separation or spacing between words. It also does not contain systematic punctuation (e.g., full stop or comma) or capitalization to indicate the overall structure. This creates an issue where search strings can be matched with any section of a text, even if it is illegible in the sentence. Therefore, some libraries provide a tokenized title to assist with discovery and retrieval. Other challenges include incomplete bibliographic information for materials published prior to the 1980s, and the complex structure of unique Thai materials such as cremation volumes.

Assigning Controlled Vocabularies and Classification

When asked which cataloging process was the most challenging, the most frequent response was creating and

applying subject headings. LCSH was the most popularly used controlled vocabulary, yet participants indicated that it was frequently insufficient to describe Thai materials. Many institutions have developed local controlled vocabularies in response to these difficulties.

Regarding classification, Dewey Decimal Classification (DDC) and LC Classification are frequently used. In places like Japan and Korea, the Nippon Classification System or Korean Decimal Classification System, derivatives of DDC, are used. As seen with controlled vocabularies, a number of institutionally based classification systems have also been created.

The Changing Library Ecosystem Related to Thai Materials

Findings from this research indicate that Thai materials are being described by people with Thai ethnicity. Yet the necessity of collecting Thai materials at all in libraries outside Thailand is being questioned due to changing budgets and collecting priorities. A number of libraries that hold a large collection of Thai materials no longer acquire new Thai materials. The National Library of Thailand has a list of libraries around the world with which they maintain material exchanges. The researchers of this paper contacted libraries on this list, and several respondents were not aware of any Thai materials within their collections.

Discussion

Applying Merton's Insiders/Outsiders doctrine, libraries outside Thailand are considered the Outsiders. However, the way that these libraries have their own interests and values is reflected in their practices with Thai materials and may indicate that they perceive themselves as Insiders. From a global perspective, knowledge about Thai people and Thai culture is considered of less of concern due to the limited socio-political power of Thailand itself. Yet for this paper, Thailand and Thai collections can be seen as an example of the representation of nondominant cultures in library collections. This relates to library topics around diverse perspectives and colonization. In a more extreme Insiders' perspective, it can be perceived that Thai materials (and materials of other nondominant groups) are treated as a peripheral domain supplementing the dominating western knowledge, which is claimed to be more conventional, universal, and neutral.

In contrast, the results indicate that the majority of Thai collections outside Thailand are dependent on the work of enthusiastic non-Thai researchers and collectors. Some collections outside Thailand may be perceived as having less value than when they are collected by Thai libraries. Donors and advocates of collections of Thai materials outside Thailand are converted Outsiders since their passion and enthusiasm can be interpreted as an act of acceptance seeking from the Thai-based scholarly community. Nonetheless, due to the influence of library users on the management of Thai collections outside Thailand, it is worth investigating further how the perceptions and experiences of users of Thai materials within and outside Thailand. Such a study would provide a more in-depth analysis on the demand of Thai materials and perhaps provide recommendations on collection management practices of Thai materials in a diverse global network environment while still respecting the identity and original value of the culture.

The practice of collecting foreign collections is common in libraries and other cultural institutions. While this study focuses on Thai collections in exile, the study findings serve as a compelling example of multidimensional perspectives on how materials of nondominant cultures are managed in a complicated global context. Findings testify to the existence and proliferation of, conceivably involuntary, social segregation and polarization in modern library collection management practices. Library services and collection management can be a part of a dialog on social segregation, prejudice, and eventually inequality. Using a customized library classification system as one of the supporting cases, Feinberg strongly argues that classification has a communicative power to persuade audiences.[68] While the variety of managing Thai materials can reflect different values and interest of the collectors and users, it may intensify the perception of dominating and submissive culture and knowledge (i.e., the Insiders/Outsiders doctrine).

Study Limitations

As with any study, there were limitations based on design and implementation. This study took a library focused approach for identifying possible collections. The researchers had language skills in Thai, English and Japanese, and could identify only a limited number of collections. There may be others available that they were unable to identify. Additionally, the research focused on publicly available library collections and did not consider private collections or collections in museums or archives. The researchers are implementing a follow up study to focus on collections as identified by Thai studies scholars that will likely include private collections. Furthermore, a large part of the data collection for this research was conducted during the 2020 global pandemic. A number of potential participants were contacted who did not respond to email inquiries due to the challenges of providing library services during a crisis. In particular, it would be valuable to get a higher level of participations by North America collections once the pandemic situation has been stabilized. If the study had been

conducted during a different time, it is very likely that participation would have been greater.

Conclusion

This study aimed to investigate the acquisition and cataloging practices of libraries outside Thailand in managing Thai materials. Applying a mixed-method approach, websites, catalogs, and LibGuides of fifty-one libraries worldwide were reviewed for their information about the characteristics of Thai materials. Purposive sampling was applied to recruit nine participants for semi-structured interviews via either in-person, by video call, or through email. The results highlight the dependency of advocate scholars in contributing Thai collections and the widely adoption of conventional Euro-American systems and standards in describing and classifying the materials. Subsequent issues regarding access and representation of Thai knowledge

systems in a broader context are also discussed. Although Thai collections of materials can be relatively small in size and considered as a niche collection in a global context, the issues of Thai materials in libraries outside Thailand can be assimilated to collections of other underrepresented communities or groups. The current global social structure of library networks and description standards play significant roles in contributing to such an issue. The result of this study exemplifies how library practices may influence the alienation or exile of nondominant cultures through acquisitions, cataloging and collection management choices.

While equal access to information has been a foundational principle of modern libraries, its interpretation and implication should be extended beyond allowance to entrance. On a global scale, library communities should engage in an in-depth analysis on how to manage foreign collections to support inclusive access and representation while still respect the values and interests of their own users.

References and Notes

1. Hope Olson and Rose Schlegl, "Standardization, Objectivity, and User Focus: A Meta-analysis of Subject Access Critiques," *Cataloging & Classification Quarterly* 32, no. 2 (2001): 61–80, https://doi.org/10.1300/J104v32n02_06.

2. Kimmo Kosonen, "The Use of Non-dominant Languages in Education in Cambodia, Thailand, and Vietnam: Two Steps Forward, One Step Back," in *Language Issues in Comparative Education*, ed. Carol Benson and Kimmo Kosonen (Rotterdam: Sense Publishers, 2013), 43, https://doi.org/10.1007/978-94-6209-218-1_3.

3. Hwa-Wai Lee, "Fragmentation of Academic Library Resources in Thai University Libraries," *International Library Review* 3, no. 2 (1971): 155–67, https://doi.org/10.1016/0020-7837(71)90050-1.

4. Hollie White and Songphan Choemprayong, "Thai Catalogers' Use and Perception of Cataloging Standards," *Cataloging & Classification Quarterly* 57, no. 7–8 (2019): 530–46, https://doi.org/10.1080/01639374.2019.1670767.

5. Henry Ginsburg, *Thai Art and Culture: Historic Manuscripts from Western Collections* (Honolulu: University of Hawaii Press, 2000), 1–138.

6. Oskar Frankfurter, "The Aims of the Society," *Journal of the Siam Society* 1 (1904): 1–6.

7. South Asian Cooperative Collection Development, "South Asia Cooperative Collection Development Workshops Statement on Cataloging and Metadata Spring 2021," accessed May 28, 2021, https://guides.lib.utexas.edu/ld.php?content_id=60138108; Committee on Research Materials on Southeast Asia, "CORMOSEA Statement On Collection Development, Access, Equity in The Time Of COVID 19," accessed

May 28, 2021, https://cormosea.files.wordpress.com/2020/09/cormosea-statement-on-collection-development-2.pdf; Council on East Asian Libraries, "Council on East Asian Libraries' (CEAL) Statement on Collection Development and Acquisition Amid the COVID-19 Pandemic: In Collaboration with the North American Coordinating Council on Japanese Library Resources (NCC) and the Society of Chinese Studies Librarians (SCSL)," accessed December 29, 2020, http://www.eastasianlib.org/newsite/wp-content/uploads/2020/08/CEAL_Statement_COVID-19_FINAL_2020.8.13.pdf.

8. Ginsburg, *Thai Art and Culture*, 1–138.

9. David Walker Ginsburg, "Studying the Neighbors: The Asian Collections," in *Remarkable Occurrences: The National Library of Australia's First 100 years 1901-2001*, ed. Peter Cochrane (Canberra: National Library of Australia, 2001): 163–81.

10. Justin McDaniel, "Illuminating Archives: Collectors and Collections in the History of Thai Manuscripts," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 2, no.1 (2017): 4, https://repository.upenn.edu/mss_sims/vol2/iss1/1.

11. McDaniel, "Illuminating Archives," 14.

12. McDaniel, "Illuminating Archives," 5.

13. Ginsburg, *Thai Art and Culture*, 1-138; Ginsburg, "Studying the Neighbors," 163-181.

14. Ginsburg, 165.

15. Ginsburg, 170.

16. McDaniel, "Illuminating Archives," 4; Ryan Wolfson-Ford, "Tai Manuscripts and Early Printed Books at the Library of Congress," *Journal of the Siam Society* 107,

no. 2 (2019): 135–52, https://so06.tci-thaijo.org/index.php/pub_jss/article/view/224703; Susanne Ryuyin Kerekes and Justin McDaniel, "Siamese Manuscript Collections in the United States," *Manuscript Studies* 2, no. 1 (2017): 202–37, https://repository.upenn.edu/mss_sims/vol2/iss1/8; Barend Jan Terwiel, "Cultural Goods and Flotsam: Early Thai Manuscripts in Germany and Those Who Collected Them," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 2, no 1 (2018): 82–105, https://repository.upenn.edu/mss_sims/vol2/iss1/3/; Larry Ashmun, "Larry and Ginny Judd SEAiT Collection" (Council on Thai Studies Annual Meeting, Indiana: Northern Illinois University, 2020), https://www.niu.edu/cots/index.shtml.

17. Kerekes and McDaniel, "Siamese Manuscript Collections," 202–37; Terwiel, "Cultural Goods and Flotsam," 82–105; Ashmun, "Larry and Ginny Judd SEAiT Collection."

18. Kerekes and McDaniel, "Siamese Manuscript Collections," 202–37

19. Jana Igunma, "Henry D. Ginsburg and the Thai Manuscripts Collection at the British Library," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 2, no. 1 (2017): 22–81, https://repository.upenn.edu/mss_sims/vol2/iss1/2/.

20. Toshiya Unebe, "Manuscripts from the Kingdom of Siam in Japan," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 2, no. 1 (2017): 151–73, https://repository.upenn.edu/mss_sims/vol2/iss1/6/.

21. Terwiel, "Cultural Goods and Flotsam," 82–105.

22. McDaniel, "Illuminating Archives," 4.

23. Andy Stephens, "Functions, Tasks and Roles of National Libraries in the 21st Century," *Alexandria: The Journal of National & International Library and Information Issues* 26, no. 2 (2016): 158.

24. Igunma, "Henry D. Ginsburg and the Thai Manuscripts Collection," 22.

25. Igunma, "Henry D. Ginsburg," 26.

26. Annabel Gallop, "Digital Access to Thai Manuscripts," *Asian and African Studies Blog*, accessed December 20, 2020, https://blogs.bl.uk/asian-and-african/thai.html.

27. Igunma, "Henry D. Ginsburg," 26.

28. British Library, "Collection Guides Thailand, Laos and Cambodia," accessed December 20, 2020, https://www.bl.uk/collection-guides/thailand-laos-and-cambodia.

29. Gallop, "Digital Access to Thai Manuscripts."

30. Library of Congress Overseas Offices, "Jakarta, Indonesia: Library of Congress Overseas Offices," accessed December 20, 2020, https://www.loc.gov/acq/ovop/jakarta/jakarta-about.html.

31. Library of Congress Overseas Offices, "Jakarta, Indonesia."

32. Gosling, "Asian Collections at the National Library of Australia," 19.

33. Gosling, "Asian Collections at the National Library of Australia," 19.

34. "Thai Collection," National Library of Australia, accessed May 28, 2021, https://www.nla.gov.au/asian/thai.

35. Tessa Morris-Suzuki, "Out of Asia: The National Library Shuts Borders and Stops Collecting on our Neighbors," *The Sydney Morning Herald*, last modified May 27, 2020, https://www.smh.com.au/national/out-of-asia-the-national-library-shuts-borders-and-stops-collecting-on-our-neighbours-20200524-p54vtp.html.

36. Tikhara, "The National Library," 215–18.

37. "About CORMOSEA," Committee on Research Materials on Southeast Asia, accessed May 28, 2021, https://cormosea.wordpress.com/about-cormosea/.

38. "Thai Collection," National Library of Australia, accessed May 28, 2021, https://www.nla.gov.au/asian/thai.

39. Igunma, "Henry D. Ginsburg," 25.

40. Ashmun, "Larry and Ginny Judd SEAiTCollection"; Walker Ginsburg, "Studying the Neighbors," 163–81; Kerekes and McDaniel, "Siamese Manuscript Collections," 202–37.

41. The bias of knowledge organization systems in libraries have been discussed in a number of publications, for instance, Clare Beghtol, "A Proposed Ethical Warrant for Global Knowledge Representation and Organization Systems," *Journal of Documentation* 58, no. 5 (2002): 507–532, https://doi.org/10.1108/00220410210441; Sandford Berman, *Prejudices and Antipathies: A Tract on the LC Subject Headings Concerning People* (Lanham, MD: Scarecrow Press, 1971); Hope Olson, "Mapping Beyond Dewey's Boundaries: Constructing Classificatory Space for Marginalized Knowledge Domains," *Library Trends* 47, no. 2 (1998): 233–54, http://hdl.handle.net/2142/8210; Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, MA: The MIT Press, 2001); Eric J. Hunter, *Classification Made Simple,*. 2nd ed. (London: Ashgate, 2006); Hollie White, "Decolonizing the Way Libraries Organize" (paper presented at IFLA WLIC 2018, Kuala Lumpur, Malaysia, 28 August 2018), http://library.ifla.org/id/eprint/2221; White and Choemprayong, "Thai Catalogers' Use."

42. Lian Tie Kho, "Cataloging and Classifying Southeast Asian Materials," in Mohammed M. Aman, ed., *Cataloging and Classification of Non-Western Material: Concerns, Issues and Practices* (Phoenix: Oryx Press, 1980), 151–234.

43. Kho, "Cataloging and Classifying," 170.

44. Kho, "Cataloging and Classifying," 171.

45. "Thai Romanization," Thammasat University Library and NECTEC, accessed May 28, 2021, http://164.115.23.167/plangsarn/.

46. Kho, "Cataloging and Classifying," 171.

47. Policy & Standards Division, Library of Congress, "Romanization Landscape," accessed December 19, 2020, www.loc.gov/catdir/cpso/romlandscape_Oct2011.html.

48. Policy & Standards Division, Library of Congress, "Romanization Landscape."

49. Cornell University Library Usability Working Group, *Thai*

Script Study (Ithaca, NY: Cornell University Library, 2020).

50. McDaniel, "Illuminating Archives," 16.
51. Cornell University Library Usability Working Group, *Thai Script Study.*
52. Cornell University Library Usability Working Group, *Thai Script Study.*
53. Bella Weinberg, "Cataloging in Non-Roman Scripts: From Radical to Mainstream Practice," in *Radical Cataloging: Essays at the Front*, ed. Keller R. Roberto (North Carolina: McFarland & Co. Inc, 2008), 29.
54. Weinberg, "Cataloging in Non-Roman Scripts," 31.
55. White and Choemprayong, "Thai Catalogers' Use," 539–41.
56. White and Choemprayong, "Thai Catalogers' Use," 539–41.
57. White and Choemprayong, "Thai Catalogers' Use," 539–41.
58. White and Choemprayong, "Thai Catalogers' Use," 539–41.
59. John W. Creswell and Vicki L. Plano Clark, "Choosing a Mixed Methods Design," in *Designing and Conducting Mixed Methods Research* (Thousand Oaks, CA: Sage, 2007), 58–88.
60. Merton, "Insiders and Outsiders," 9-47.
61. Elfreda A. Chatman, "Life in a Small World: Applicability of Gratification Theory to Information-Seeking Behavior," *Journal of the American Society for Information Science* 42, no. 6 (1991): 438–49, https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199107)42:6%3C438::AID-ASI6%3E3.0.CO;2-B.
62. Merton, "Insiders and Outsiders," 9–47.
63. Merton, "Insiders and Outsiders," 15.
64. Merton, "Insiders and Outsiders," 15.
65. Chatman, "Life in a Small World," 438–49; Songphan Choemprayong, "The Transition of Worldviews: Collective Information Behavior during the 2006 Thai Coup D'état" (PhD diss., University of North Carolina at Chapel Hill, 2010), https://doi.org/10.17615/gzwa-q521.
66. Jennifer Watling Neal et al., "Expediting the Analysis of Qualitative Data in Evaluation: A Procedure for the Rapid Identification of Themes from Audio Recordings (RITA)," *American Journal of Evaluation* 36, no. 1 (2015): 118–32, https://doi.org/10.1177/1098214014536601.
67. Non-Thai librarian participant, February 13, 2020.
68. Melanie Feinberg, "Classification as Communication: Properties and Design" (PhD diss., University of Washington, 2008), accessed January 15, 2021, https://ils.unc.edu/~mfeinber/Feinberg%20dissertation.pdf.

Appendix A: Interview Questions

In general, we would like to learn about your experience in cataloging and maintaining Thai materials. These are our initial questions. After receiving your responses, we may have additional questions to follow.

1. What type(s) of training and qualifications did you have to obtain to get to do this job?
2. What type of materials do you catalog most frequently?
3. Could you describe the process you would go through to add a Thai material to your library catalog?
   3.1 What is your original cataloging workflow like?
   3.2 What is your copy-cataloging workflow like?
4. What type of cataloging resources do you use/consult at your job?
   4.1 For subject headings?
   4.2 For call numbers?
   4.3 For descriptions?
5. Among the following tasks, which part of cataloging is the most challenging for Thai materials? and why?
   5.1 creating subject headings is the hardest task
   5.2 creating call numbers is the hardest task
   5.3 creating descriptions is the hardest task
6. Do you use local systems to organize your materials? If so, could you explain about these systems?
7. What resources and networks do you consult when you have questions about cataloguing Thai materials?
8. What institutions do you consider are authorities for Thai bibliographic data?

Further questions address acquisitions and collecting approaches based off of information included in the answers above.

Exploration of Subject Representation and Support of Linked Data in Recently Created Library Metadata

Examination of Most Widely Held WorldCat Bibliographic Records

Vyacheslav Zavalin, Oksana L. Zavalina, and Shawne D. Miksa

This paper presents results of the examination of subject representation in the most recently created library metadata records. The bibliographic records were collected from the WorldCat global database. The records were created in 2020 according to the latest version of Resource Description and Access (RDA) and MARC 21Format for Bibliographic Data. A purposive sample of the records with the widest reach—as expressed in the highest number of holdings and the highest level of editing made by multiple institutions—was selected for in-depth content analysis. The level and patterns of application were analyzed for all subject representation data elements (record fields and subfields), specifically for those that were Linked-Data-enabling. The study examined the level and patterns of application of subject controlled vocabularies. Co-occurrences between various subject representation data elements and between subject controlled vocabularies within the records were explored.

Vyacheslav Zavalin (vzavalin@twu.edu) is an Assistant Professor in the School of Library & Information Studies at Texas Woman's University. **Oksana L. Zavalina** (oksana.zavalina@unt.edu) is an Associate Professor in the Department of Information Science at the University of North Texas. **Shawne D. Miksa** (shawne.miksa@unt.edu) is an Associate Professor in the Department of Information Science at the University of North Texas.

Helping users to satisfy their information needs and obtain needed information resources is the top priority in the field of library and information science. The representation of information objects through metadata is a key activity of libraries, archives and museums that is necessary to provide access to recorded knowledge held by those institutions. Several types of metadata records are used by these communities. Metadata records that represent information objects are commonly referred to as bibliographic records. The most common data traditionally included in bibliographic records are titles and subjects of works, plus the names of their creators.

In the current information environment, the amount of generated data and published information continues to rapidly increase and is often referred to as an information explosion, resulting in information overload.[1] As resource discovery by title or creator of an information object is seriously limited by this information

overload, resource discovery by subject becomes even more important.[2] This places an increasing emphasis on the functionality of subject metadata—the parts of bibliographic records that represent the aboutness of information objects.[3] The creation of subject metadata relies not only on analysis of aboutness, but also on examination of relationships among topics, form, and genre in the context of the intended audience and possible uses of information objects.[4]

MARC 21 Format for Bibliographic Data (MARC 21) is currently the dominant cataloging encoding format for description of information objects and the exchange of metadata among databases. Development of Linked Data potentially improves discoverability of information through metadata records, including subject access through subject metadata. The new BIBFRAME standard is developed with Linked Data functionality support in mind, and will eventually replace MARC 21. Millions of existing MARC 21 records that collectively represent and provide access to the vast body of recorded knowledge will need to be reformatted or converted from MARC 21 to BIBFRAME. Due to the sheer volume of that conversion task, the reformatting will need to be automated. As the output quality in automatic conversion processes relies greatly on the input quality, to ensure the conversion produces meaningful and functional results, the input metadata (data values in the fields of MARC 21 bibliographic records) needs to support that functionality. However, it is unclear as to what extent the Linked Data functionalities can be realized when the records are converted automatically from MARC 21 to BIBFRAME. This paper reports selected results of the exploratory study that sought answers to this question, with a focus on the subject representation in library bibliographic metadata records.

Literature Review

Bibliographic records are created according to several types of standards. Currently, the most widely used data content standard in the library community is Resource Description and Access (RDA).[5] The prevailing data encoding and transmission standards are the well-established MARC format, and the more recent alternative, Bibliographic Framework Initiative (BIBFRAME)—both of which are metadata element sets.[6] The data value standards include controlled vocabularies (e.g., thesauri, lists of subject terms and codes, etc.) and classification schemes.

RDA is an international standard that began to be developed in 2005 and was officially implemented by the Library of Congress (LC) in 2013. It was initially planned as a third major revision to the Anglo-American Cataloguing Rules (AACR), and evolved into a new standard with a substantially different conceptual base. RDA continues to evolve to meet user needs. The recently completed 2020 revision of RDA (3R) has not yet been widely adopted by metadata practitioners due to usability issues, and is expected to be adopted in 2023; most catalogers currently rely on the April 2018 revision.[7] Development of RDA is informed by the ideas of the Semantic Web, which seeks to connect pieces of information in a logical way that is understandable and can be processed by machines to improve information retrieval.[8] This way of connecting information is called Linked Data. One of the most important steps in ensuring its validity is the inclusion of unique Uniform Resource Identifiers (URIs) that link to openly available information on the entity in question and related entities.

BIBFRAME builds upon application of Linked Data principles to bibliographic metadata and is projected to replace the MARC 21 standard. BIBFRAME metadata record creation tools are being developed and explored by the early adopters, and software companies are starting to incorporate them into the integrated library systems such as for example Ex Libris' Alma. Until these tools become mainstream (a process that will take years, if not decades), most newly created records will follow the MARC 21 standard. Currently, hundreds existing MARC 21 bibliographic records collectively provide access to the body of recorded knowledge, and MARC 21 maintains its importance as an encoding standard.[9] Furthermore, the MARC 21 bibliographic metadata element set constantly evolves as new fields and subfields are added to support the Linked-Data-related and other RDA requirements.

Since the beginning of RDA's development, several new subject representation data elements have been added to the MARC 21 bibliographic element set to improve functionality and to support Linked Data. As part of these revisions, MARC 21 has been expanded to include new subfields in variable fields that enable the inclusion of URIs into bibliographic records. The 650 (Subject Added Entry—Topical Term) is one of the fields for which the subfield $4 Relationship was initially added to MARC 21 Bibliographic Standard in 2007. This subfield was later renamed as Relator Code and redefined in 2017 to include URIs as data values. Additionally, in 2017 the subfield $1 Real World Object URI, was added to MARC 21 Bibliographic Format for this use in several fields. The subfield $0 Authority record control number that had been part of the standard since 2005 was emphasized after 2013 as the way to link bibliographic records to authority records. The library community is working to improve the Linked Data functionality of existing MARC 21 bibliographic records by enriching them with URIs, including subject metadata fields.[10] Recent publications include reports of converting the Program for Cooperative Cataloging (PCC) MARC 21 records to BIBFRAME, comparative evaluations of Linked Data ontologies and data models as they apply to MARC,

and discussion of the future of authority control in libraries in the Linked Data environment.[11] Zeng and Mayr shared results of their review of how existing knowledge organization systems used in library metadata (including subject controlled vocabularies) can transition to become Linked Open Data.[12]

The principles of building subject controlled vocabularies have been developed and refined in communities of information professionals for many years, with the first major controlled vocabularies appearing in the nineteenth century. In the twenty-first century, construction of subject controlled vocabularies is guided by the International Standard for Thesauri and Interoperability with Other Vocabularies, currently in version 1.4.[13] More than three hundred controlled vocabularies for subject representation have been developed and are maintained worldwide.[14] Some are multilingual, such as the French-English Répertoire de vedettes-matière (RVM), which is used for verbal subject representation in Canada. Subject controlled vocabularies are often developed by national libraries or archives such as the Gemeinsame Normdatei, which is developed by the German National Library or the US Library of Congress Subject Headings (LCSH). Faceted Application of Subject Terminology (FAST) is a derivative controlled vocabulary that relies on LCSH, and provides an added level of functionality by splitting LCSH subject strings into facets. Book Industry Study Group (BISAC) subject headings are another subject controlled vocabulary of general applicability intended for use by publishers and bookstores.[15] A number of subject controlled vocabularies focus on specific domains: one example is the Medical Subject Headings (MeSH) used to represent works originating in the biomedical knowledge domain.

Controlled vocabularies for subject representation include the broad classifications that cover the entirety of human knowledge, such as DDC, Library of Congress Classification (LCC), Universal Decimal Classification (UDC), and classification systems that focus on specific knowledge domains (e.g., National Agricultural Library classification for agricultural materials, Government of Canada Classification for government publications, etc.). Some subject controlled vocabularies combine verbal and nonverbal subject representation—one example is the Chinese Classified Thesaurus (Zhong guo fen lei zhu ti ci biao).[16]

Tools and technologies that enable automatic and semi-automatic generation of metadata from the full-text of textual information objects use indexing to help expand access to information. Despite its advantages, full-text indexing cannot provide the same level of access as subject representation with controlled vocabularies. Research shows deficiencies for information retrieval, for example, in representing foreign language materials.[17] They are also not useful for creating metadata for non-textual information objects (e.g., works of music, visual art, photographs, etc.).

Smith-Yoshimura et al. emphasized providing controlled-vocabulary subject access in the creation of MARC 21 records: "The number of full-text documents available on the Web will substantially increase over the next few years, and the need for surrogate 'descriptive metadata' will decrease. Focus instead on the authorized names, classifications, and controlled vocabularies that key word searching of full-text will not provide."[18] As of 2015, automated indexing tools were considered as not yet sufficiently developed for full-scale implementation by the library community in a meaningful way.[19] However, a promising new multilingual automated subject indexing tool Annif has been developed and used by early adopters in the international library community for digital collections in 2017-2020.[20]

Studies of MARC 21 library metadata typically draw datasets for analysis from large databases such as the LC's catalog or OCLC's WorldCat. The advantage of the WorldCat database compared to a single library catalog (even as large as the LC's) as a source of data for studying worldwide cataloging practices is its heterogeneity and global impact. Since 1998, WorldCat has been the major centralized shared database of bibliographic records created and edited collaboratively by the international library community. The WorldCat database is widely used in fulfilling interlibrary loan (ILL) requests.[21] It is also a major tool used in cooperative cataloging worldwide: when a library or other institutional member of WorldCat adds an item to its collection, it either

- submits a new bibliographic record to the database; or
- if the record is already in WorldCat, uses the existing record is used as is or edits the master record and updates the holdings information by indicating that it has the item in its collection.

In both copy and original cataloging, bibliographic records are also added to the institution's online catalog. As this paper was being finalized, a new record was added to the WorldCat database every second, and contained more than 516 million metadata records representing information objects in 483 languages.[22]

Several studies of MARC 21 bibliographic metadata have examined subject representation in library metadata records. Almost all were completed pre-RDA, meaning that they were conducted before new subject data elements (fields and subfields, including Linked-Data enabling ones) were added to the MARC 21 standard and applied to bibliographic records. Furthermore, most of the studies conducted since 2000 did not focus on subject metadata. Relevant findings from these studies are reviewed below.

Moen and colleagues found that the MARC 21 Field 600 Subject Added Entry—Personal Name is the one

most frequently occurring subject-related field in MARC 21 bibliographic records.[23] Eklund et al. observed that the MARC field 655 Index Term—Genre/Form was present in only 5 percent of records for sound recordings.[24] Mayernik observed that the MARC field 650 Subject Added Entry—Topical Term appeared in 66 percent of records and exhibited the largest average number of occurrences (1.84 per record), and that other subject representation fields–050 Library of Congress (LCC) Call Number, 043 Geographic Area Code, and 082 Dewey Decimal Classification (DCC) Number–were among the most frequently occurring MARC 21 fields.[25] Smith-Yoshimura et al. noted that four subject metadata fields were among the top twenty-two most frequently occurring MARC 21 fields: 650 (46 percent of records), 050 (20 percent), 043 (19 percent), and 082 (14 percent). Smith-Yoshimura's team also separately examined the application of fields recently added to the MARC 21 standard as of the time of their analysis and found subject metadata fields 648 Subject Added Entry-Chronological Term and 662 Subject Added Entry—Hierarchical Place Name to be used in under 0.1 percent of records.[26] Moen and Benardino examined 400,000 MARC 21 bibliographic records and observed 122 different MARC 21 subject metadata subfields (e.g., 650 $v Subject Added Entry Topical Term—form subdivision, 651 $y Subject Added Entry Geographic Term—chronological subdivision, etc.).[27]

Taylor and Simpson compared LC's Cataloging-In-Publication records with other bibliographic records, and found mistakes and omissions in subject headings, geographic area codes, DDC and LCC classification codes.[28] In her analysis of records from two databases, Intner discovered a lack of subject headings or classification numbers in records.[29] In the meta-analysis of subject search in online catalogs, Larson summarized improvements in subject representation in bibliographic records that had been proposed by researchers. These suggestions included assigning more LCSH headings per record, supplementing them with terms from specialized thesauri (e.g., MeSH), providing more specific class notations, assigning additional class numbers to represent multiple facets of a work, etc.[30] Hoffman examined the practice of facilitating subject access through the creation of individual bibliographic records with more specific subject headings for each work aggregated in a multi-work item instead of assigning more general subject headings in a single record describing the whole item.[31]

A more recent study by Zavalina, Shakeri, and Kizhakkethil examined RDA-based MARC 21 bibliographic records to determine the quantitative patterns of change between 2013 and 2015 in the application of subject metadata fields in video recording records. That study found a slight increase in the use of Linked-Data-enabling subfields, but reported low overall level of their application.

The authors observed the overall trend toward an increase in the average number of subject headings per record.[32] The replication of these analyses in 2021 to compare the 2020 versions of the same records representing video recordings to their 2015 versions demonstrated an increase in the trend for addition of further subject fields and subfields to existing records. However, that study found some decrease in the level of application of Linked-Data-supporting subfield $0 Authority record control number or standard number in field and/or in the average number of instances of this subfield per record that includes the field in five subject metadata fields.[33] Zavalin examined the application of subject and genre controlled vocabularies in a sample of 688 WorldCat bibliographic records contributed by the LC's Children's and Young Adults' Cataloging (CYAC) Program between 2014 and 2020. The author observed the use of twenty genre and eighteen subject controlled vocabularies.[34]

None of the previous studies of subject representation in MARC 21 bibliographic records examined the use of the co-occurrences of data elements and controlled vocabularies within a record. Additionally, no published studies of MARC content designation focused on the examination of subject metadata in records created after the major revision of RDA in 2018 and following the latest addition of a new subject field to the MARC 21 Bibliographic format: the field 688 Subject Added Entry—Type of Entity Unspecified in 2019. This study seeks to be one of the first to analyze records following these and other recent changes to library metadata standards to develop an understanding of the current level and patterns in subject representation, including support for Linked Data, as evidenced in a set of MARC 21 bibliographic records that are included in a large number of library catalogs.

Method

This study used the content analysis of the recently created RDA-based MARC 21 bibliographic metadata records in the WorldCat database. A purposive sample of 100 MARC 21 records that were created in 2020, and thus were expected to follow the most recent versions of RDA and MARC 21 standards, was selected, based on several major criteria. First, records that represent information objects that were held in at least 500 library collections at the time of data collection in May 2020 were targeted. Records with such a high level of holdings have the greatest impact on access to information (including subject access) in library collections. Also, regardless of age, these widely held records are typically edited more than once by multiple institutions since their creation. Second, records were selected with the highest overall quality as indicated by the "full level of encoding" code in the ELvl subfield of the fixed field. Specifically,

the authors targeted those with the code blank as "the most complete MARC record[s] created from an inspection of the material" or those with the code "I," indicating the next most complete full-encoding level.[35]

The authors used MarcEdit, a metadata manipulation and editing software suite, to collect the data using the Z39.50 client-server protocol developed for searching and retrieving information from remote databases through Transmission Control Protocol/Internet Protocol (TCP/IP) supporting networks by LC's Maintenance Agency. The application of criteria discussed above and the deduplication of the list of records matching these criteria resulted in a set of one hundred unique metadata records.

An in-depth content analysis of these metadata records was performed. The study examined general characteristics such as types and languages of materials represented, types and locations of institutions that created records, language of cataloging, etc. The focus was on the levels of application of various subject fields and selected subfields, including Linked-Data-supporting data elements, co-occurrence between subject data elements intended for the same type of information within a record, and levels of application of subject controlled vocabularies and co-occurrence within records.

A common limitation of content analysis is researcher bias, which is normally alleviated by using detailed coding manuals, coding by multiple coders, and subsequent evaluation of the intercoder agreement. However, the design of this study bypassed researcher bias because only objective (i.e., mostly numeric and binary) characteristics and measures were assessed. No subjective evaluations (e.g., those regarding the accuracy of subject metadata) were included.

Findings

General Characteristics of Records in the Sample

Based on the holdings data attached to the records collected from WorldCat, the number of institutions that included the analyzed records in their catalogs at the time of data collection ranged between 577 and 1,514. The material types represented by the collected records were distributed as follows: books (83 percent, including regular print books, 76 percent, online books 3 percent and large print books 4 percent), visual materials (13 percent, including online materials 1 percent), sound recordings (3 percent), and continuing resources (1 percent). Forty percent of the records were created as part of LC's Copy Cataloging program (lccopycat), with headings "verified with the relevant authority file, except those subject headings not from LCSH."[36] An

additional 8 percent were created under the Program for Cooperative Cataloging's (PCC) auspices, which means that "subject headings are checked for authorized forms and combinations supported by the relevant authority."[37]

The sampling approach did not limit data collection to any specific language of items represented by records or any specific language of cataloging. However, analysis demonstrated that all records in this sample of the most widely held WorldCat bibliographic records with the highest self-identified completeness represented only English-language materials. The records were created by thirty-one institutions from six countries, with English as the language of cataloging: Australia, Canada, Hong Kong, New Zealand, the United Kingdom, and the US. The records were contributed to WorldCat by academic libraries (e.g., University of Hong Kong's library), school libraries (e.g., Anchorage school district library in Alaska), public libraries (e.g., Winnipeg Public Library), state/national libraries (e.g., Libraries Australia), federal/national government agencies (e.g., US National Library of Medicine), associations/foundations (e.g., Libraries Horowhenua in New Zealand), and vendors and other corporate/business organizations (e.g., Baker & Taylor). The number of records contributed by each institution ranged from one to twenty-nine, with 3.225 on average.

Application of Subject Representation Data Elements

Table 1 shows the level of application of observed MARC 21 subject representation metadata fields. The dataset contained a total of eighteen MARC 21 bibliographic metadata fields for subject representation (see table 1). At the time of data collection, all but one of these subject metadata fields, field 043 Geographic Area Code, were repeatable, meaning that more than one instance of a field could be included in a bibliographic record. However, with the December 2020 publication of Update no. 31 to MARC 21 Format for Bibliographic Data, the 043 also became repeatable.[38]

As shown in table 1, only the 650 Subject Added Entry—Topical Term field was present in all records. The number of instances of this field varied between two and forty-six instances. Three other fields appeared in 98 percent of the records. This included two fields that provide classification data and one that represents genre: 050 Library of Congress Call Number, 082 Dewey Decimal Classification Number, and 655 Index Term—Genre/Form. The level of application of the remaining fourteen subject representation fields in MARC 21 bibliographic records ranged widely between 1 percent (fields 080 Universal Decimal Classification Number, 092 Locally Assigned Dewey Call Number, and 654 Subject Added Entry—Faceted Topical Terms) and 59 percent of records (field 651 Subject Added Entry—Geographic Name). The average number of

Table 1. Level of application of observed MARC 21 subject representation metadata fields

Field	% of Records with 1+ Instance	Ave. No. of Instances per Record	Median	Mode	Range	Variance	Standard Deviation
043 Geographic Area Code	53	1.0000	1	0	0-1	0.2516	0.5016
050 Library of Congress Call Number	98	1.0204	1	1	0-2	0.0404	0.2010
055 Classification Numbers Assigned in Canada	5	1.0000	0	0	0-1	0.0480	0.2190
060 National Library of Medicine Call Number	3	1.0000	0	0	0-1	0.0294	0.1714
072 Subject Category Code	2	1.5000	0	0	0-2	0.0496	0.2227
080 Universal Decimal Classification Number	1	1.0000	0	0	0-1	0.0100	0.1000
082 Dewey Decimal Classification Number	98	1.0204	1	1	0-2	0.0404	0.2010
084 Other Classification Number	12	1.0833	0	0	0-2	0.1344	0.3667
092 Locally Assigned Dewey Call Number	1	1.0000	0	0	0-1	0.0100	0.1000
600 Subject Added Entry-Personal Name	28	2.3571	0	0	0-6	1.5600	1.2490
610 Subject Added Entry-Corporate Name	7	2.0000	0	0	0-2	0.2630	0.5129
611 Subject Added Entry-Meeting Name	4	1.2500	0	0	0-2	0.0682	0.2611
647 Subject Added Entry-Named Event	6	1.1667	0	0	0-2	0.0860	0.2932
648 Subject Added Entry-Chronological Term	16	1.0000	0	0	0-1	0.1358	0.3685
650 Subject Added Entry-Topical Term	100	13.3500	12	12	2-46	56.3914	7.5094
651 Subject Added Entry-Geographic Name	59	2.5424	1	0	0-8	2.8586	1.6907
654 Subject Added Entry-Faceted Topical Terms	1	3.0000	0	0	0-3	0.0900	0.3000
655 Index Term-Genre/Form	98	6.9286	7	8	0-19	14.5110	3.8093

**Table 2.** Total number of various subject metadata fields and field instances per record

	Mean	Median	Mode	Range	Variance	Standard Deviation
No. of Different Subject Fields per Record	5.99	6	6	3-10	2.51505	1.585891
Total No. of Instances of All Subject Fields per Record	25.7	26	27	5-68	80.2929	8.960632

instances of fields 650 (13.35) and 655 (6.93) was the highest. Subject representation fields 651, 600 (Subject Added Entry—Personal Name), and 610 (Subject Added Entry— Corporate Name) appeared in two or more instances per record on average (2.54, 2.35, and 2.00). The highest level of variability as expressed in standard deviation of 7.51 was observed in field 650. Also, relatively high standard deviation between 1.25 and 3.81 was observed in three additional fields: 655 (Index Term—Genre/Form), 651, and 600. In the remaining fourteen fields, the standard deviation was below 0.6, which indicates consistent levels of application of a field across the records in the purposive sample.

Table 2 shows that on average, a total of six various subject fields appeared in records, with the range of three to ten. The total number of instances of all subject fields combined ranged much more substantially from five to sixty-eight per record. The central tendency measures— mean, median, and mode—for the number of instances of all subject fields combined per record were between 25.7

and 27. The analysis demonstrated high variability for the total number of subject field instances (variance of 80.29 and standard deviation of 8.96) and relatively moderate variability for the number of subject fields (variance of 2.52 and standard deviation of 1.59).

Ninety-eight percent of records in the sample included one or more instances of subfield $0 Authority Record Control Number or Standard Number, which is considered as the most important Linked-Data-enabling MARC 21 subfield.[39] A total of 778 instances of this subfield, as shown in figure 1, appeared in seven subject metadata fields: 600, 610, 611, 647, 650, 651, and 655. Almost 85 percent total of all instances of subfield $0 in subject representation fields occurred in the two most widely applied fields: 650 (48.2 percent) and 655 (36.89 percent). All instances of subfield $0 Authority Record Control Number or Standard Number observed in subject representation metadata fields included data values expressed as literals as opposed to URIs. The Linked-Data-enabling subfield $0 appeared only in the instances of 6XX
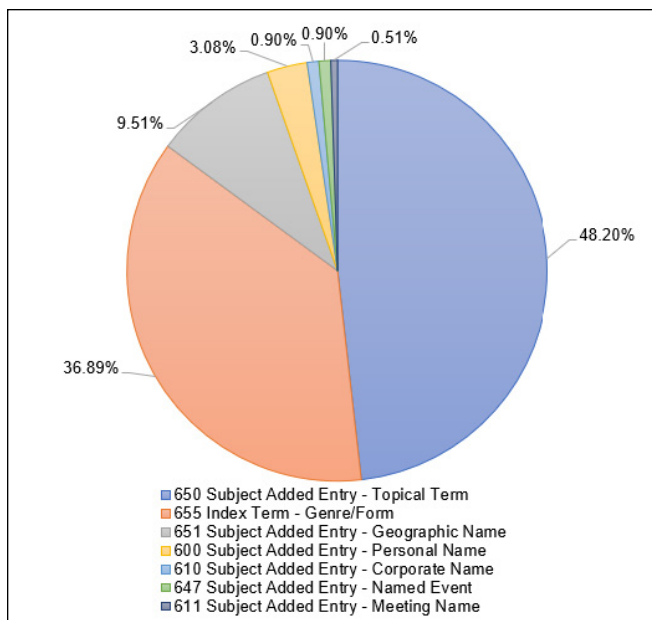
Figure 1. Distribution of Linked-Data-enabling subfield $0 Authority Record Control Number or Standard Number (% of all instances of subfield $0 observed in the sample)

fields that contained FAST headings (with exception of when FAST headings were used in field 648 Subject Added Entry-Chronological Term where it was not observed). The $0 subfield was not observed in any instances of 6XX fields that included terms from the other subject controlled vocabularies: LCSH, LC Children's Subject Headings (CSH), Medical Subject Headings (MeSH), Répertoire de Vedettes-Matière (RVM), BISAC Subject Headings, Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc. (GSAFD), Library of Congress Genre and Form Thesaurus (LCGFT), Sears subject headings, Gemeinsame Normdatei (GND), and GOO-Trefwoorden Thesaurus by Koninklijke Bibliotheek in the Netherlands (GTT). No record in the sample included any of the two other Linked-Data-enabling subfields in subject metadata fields: $1 Real World Object URI and $4 Relationship.

The application of three additional subfields in subject representation fields—043 $a Geographic Area Code and subfields $z Geographic

Subdivision and $y Chronological Subdivision in 6XX fields—was examined and compared to the application of other MARC 21 subject metadata elements intended for representing chronological and geographical aboutness of information objects. Table 3 presents the overall levels of application of these subfields. The largest number of instances was observed for $z Geographic Subdivision: it occurred seventy-two times in a total of thirty-three records in the sample. Subfield 043 $a Geographic Area Code occurred in a larger proportion of records (53 percent) but in a smaller overall number of instances (sixty-two). Subfield $y Chronological Subdivision was the least frequently used: sixteen instances total were observed in 9 percent of records. The average number of instances was the lowest for 043 $a (1.1698) and the highest for 6XX $z (2.1818). The mode number of instances was zero for all three subfields, and only the 043 $a exhibited a median number of instances above zero. The highest variability in the level of application was observed for 6XX $z.

Application of Controlled Vocabularies

LCSH was used most often to represent subject content of information resources represented by the records. It was observed in six MARC 21 subject representation metadata fields: 600, 610, 611, 650, 651, and 655. Table 4 provides the level of application of LCSH controlled vocabulary in these fields. The highest level of use of the LCSH occurred in fields 650, 600, and 651: an average of 4.05, 1.44, and 1.13 instances of the field respectively. However, the median and mode number of instances of a field with data values

Table 3. Level of application of three subject metadata subfields

Field	% of Records with 1+ Instance	Ave. No. of Instances per Record	Median	Mode	Range	Variance	Standard Deviation
043 $a	53%	1.169811	1	0	0-4	0.693112	0.480404
6XX $z	33%	2.181818	0	0	0-9	1.484465	2.203636
6XX $y	9%	1.777778	0	0	0-7	0.76171	0.580202

Table 4. Level of application of the LCSH controlled vocabulary

Subject Representation Field	% of Records with 1+ Instance	Ave. No. of Instances per Record	Median	Mode	Range	Variance	Standard Deviation
600	27%	1.4444	0	0	0-3	0.5635	0.7507
610	7%	1.0000	0	0	0-1	0.0658	0.2564
611	1%	1.0000	0	0	0-1	0.0100	0.1000
650	99%	4.0505	4	3	0-15	5.0403	2.2451
651	36%	1.1389	0	0	0-2	0.3453	0.5877
655	34	1.0294	0	0	0-2	0.25	0.5000

from LCSH, as well as variance and standard deviation, were under 0.76 for all except field 650.

Figure 2 shows the application of other subject controlled vocabularies in 6XX fields. A total of ten other subject controlled vocabularies was observed. The two most widely applied (i.e., found in over 90 percent of records) non-LCSH controlled vocabularies were the Faceted Application of Subject Terminology (FAST) and the Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT). Terms from four others—BISAC Subject Headings List (BISAC), Sears List of Subject Headings (SEARS), Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc. (GSAFD), and Children's Subject Headings (CSH)—were found in between 37 percent and 72 percent of records. The Medical Subject Headings

Figure 2. Level of application of other subject controlled vocabularies in 6XX fields

(MeSH), Gemeinsame Normdatei (GND), Répertoire de vedettes-matière (RVM), and GOO-Trefwoorden Thesaurus (GTT) were used much less often: between 1 percent to 4 percent of records. Finally, 12 percent of records included one or more instances of 6XX field(s) with the controlled vocabulary not specified.

Two classification fields—072 Subject Category Code and 084 Other Classification Number—were found to contain source code "bisacsh" indicating the terms were drawn from BISAC Subject Headings List. A total of six instances of field 072 in three records (100 percent of records with that field in the sample) and a total of twelve instances of field 084 in eleven records (91.67 percent of records with that field in the sample) included BISAC terms. Between one and two instances of three other controlled vocabularies used in these fields were observed for: (1) Book Industry Communication Standard Subject Categories (indicated by code "bicssc"), (2) Nederlandse Basisclassificatie (Dutch Basic Classification Code; indicated by code "bcl"), and (3) Elizabeth M. Moys Classification and Thesaurus for Legal Materials (indicated by code "moys").

Co-occurrence of Subject Data Elements and Controlled Vocabularies

As previously shown in table 2, between three and ten different subject fields were observed in each record, with an average of six. Certain pairs of subject metadata elements
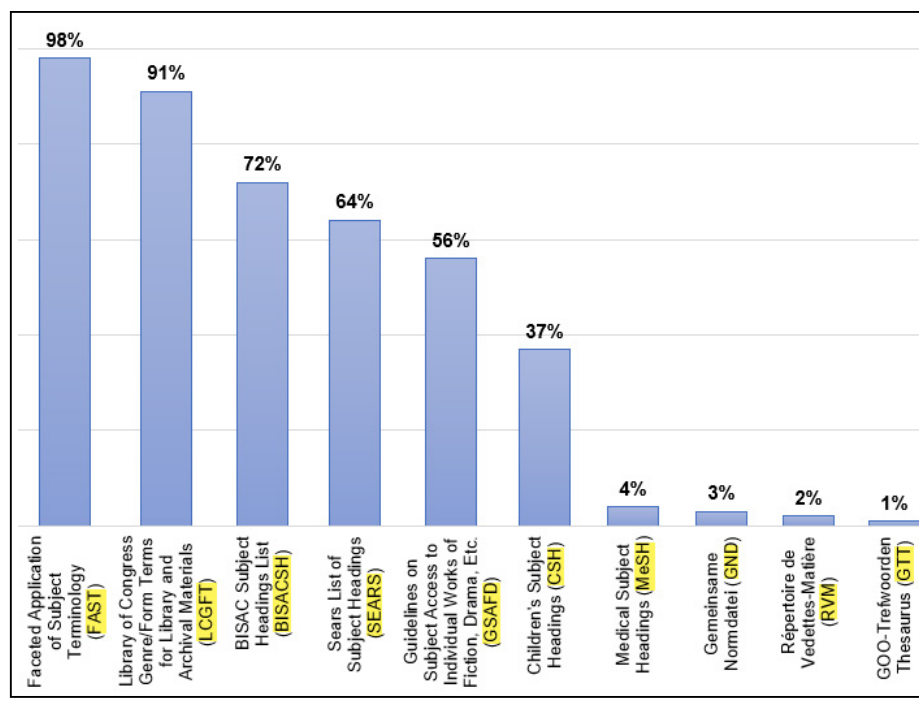
(fields/subfields combinations) providing similar or related types of information often co-occurred. Most (94 percent) records included two classification fields: 050 Library of Congress Call Number and 082 Dewey Decimal Classification Number. The co-occurrence between these two fields was the second highest, after the pair of fields 650 Subject Added Entry-Topical Term and 655 Index Term-Genre/Form that occurred together in 98 percent of records. Analysis also revealed noticeable levels of co-occurrence of

- field 648 Subject Added Entry-Chronological Term and subfield $d Date of meeting or treaty signing;
- field 611 Subject Added Entry-Meeting Name (84 percent of records in the sample);
- field 648 and subfield $y Chronological subdivision in 6XX fields (83 percent);
- field 043 Geographic Area Code and subfield $z Geographic subdivision in 6XX fields (43 percent); and
- fields 043 Geographic Area Code and 651 Subject Added Entry-Geographic Name (39 percent).

Co-occurrences between other subject fields (e.g., 650 and 651, 600 and 610, etc.) and pairs of classification fields other than 050 and 082 was much lower.

Certain pairs of subject controlled vocabularies were used together in the same records. Figure 3 presents these findings for most frequently co-occurring controlled vocabularies. LCSH and FAST were found most often together
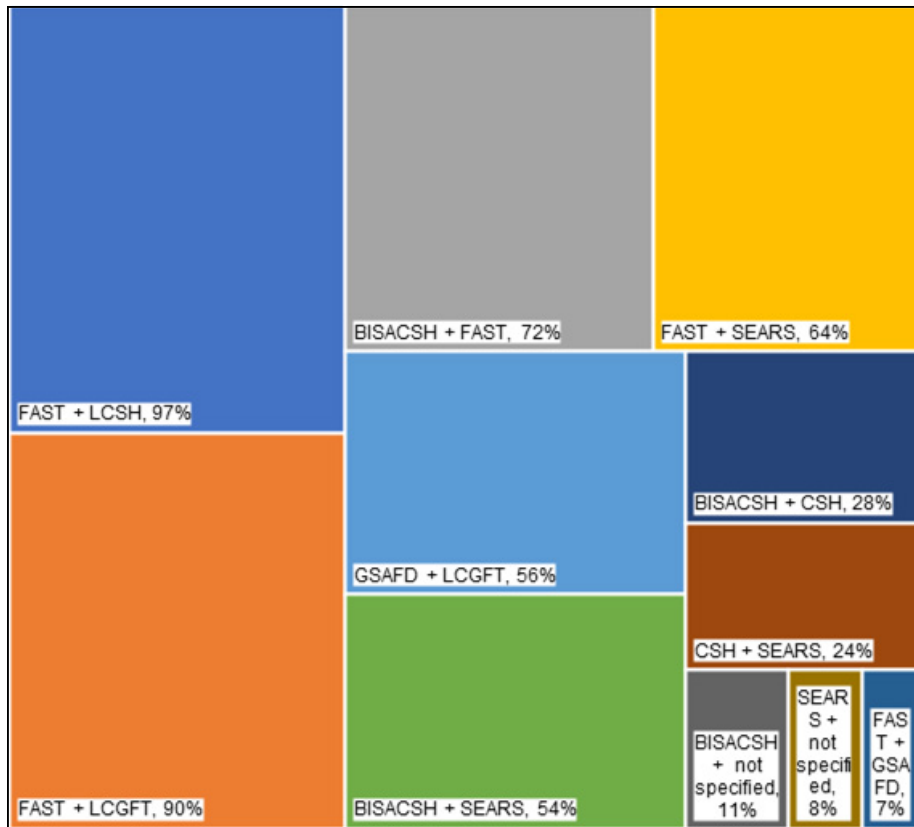
Figure 3. Co-occurrence of subject controlled vocabularies within the same records

(97 percent of records). Both FAST and LCGFT terms were included in 90 percent of records. Four additional pairs of controlled vocabularies co-occurred in more than 50 percent of records overall: FAST and BISAC subject headings (72 percent), FAST and Sears subject headings (64 percent), LCGFT and GSAFD genre headings (56 percent), and Sears and BISAC subject headings (54 percent). Although not shown in figure 2, it is worth noting that the lowest level of co-occurrence (1 percent of records in the sample) was observed for the terms from MeSH and BISAC subject headings, MeSH and Sears subject headings, Répertoire de Vedettes-Matière (RVM) and BISAC subject headings, and RVM and Sears subject headings. No co-occurrences were observed between MeSH and other controlled vocabularies beyond BISAC and Sears, or between RVM and other controlled vocabularies beyond BISAC and Sears.

Discussion

Patrick Wilson posited that

> the sense of a position in an organizational scheme is given by the rules of assignment and by what we

can deduce from those rules. When position is assigned on the basis of identification of the subject and selection of the most closely fitting position, whatever sense we have of positions depends on what we know about how it is decided what the subject of a writing is, hence what it means to say of a writing that its subject is this or that.[40]

This study is the first to provide insight into the patterns of application of subject representation in the MARC 21 records created by libraries worldwide using the latest revisions of RDA and MARC 21 metadata element set to facilitate increased Linked-Data functionality of library metadata. The findings indicate that the available MARC 21 content designation intended to support this functionality is not currently being used to its full capacity. Only one of the three Linked-Data-enabling subfields was observed in the analyzed records, with URIs for the terms from just one of the controlled vocabularies (FAST). This omission means that when MARC 21 records are converted to BIBFRAME 2.0, URIs for most controlled-vocabulary terms would not be included, and for subject representation other than that with FAST (based on LCSH), records would mostly rely on literal data values (strings of characters) that have no Linked Data power.

Overall, the findings demonstrate that subject representation has substantially increased in extent and variety compared to MARC 21 metadata created earlier and analyzed in previous studies conducted between 2003 to 2010.[41] There is an especially noticeable increase in the level of application observed for fields 650 Subject Added Entry-Topical Term, 655 Index Term-Genre/Form, and 651 Subject Added Entry-Geographic Name. The practice of enriching records by adding non-LCSH subject terms from a variety of controlled vocabularies of topical terms observed in the records analyzed in this study significantly expands subject representation in records, and, if accompanied with Linked-Data-enabling metadata elements, greatly increases functionality of bibliographic records in supporting the Explore user task as defined in the *Library Reference Model* (LRM).[42]

Eighteen of thirty-seven subject metadata fields defined in the latest version of MARC 21 metadata element

set were found in the records analyzed for this study. Both the average number of various subject fields included in records (5.99), and the average number of instances of these subject fields (25.7) are substantially higher in RDA compliant records created in 2020 than was observed in the previous studies of pre-RDA MARC 21 bibliographic records. However, it is important to note that several new subject metadata fields have been added to the MARC 21 metadata element set since the time these pre-RDA studies were completed. These include fields 083, 085, 086, and 688 that were not observed in the present study, plus fields 084 and 647, which were observed. Also, surprisingly, even within this purposive sample of the most complete cataloging records based on the full-level cataloging standard followed, a substantial variability was observed (as measured through variance and standard deviation indicators) in the application of several key subject metadata fields, including the MARC 650 Subject Added Entry-Topical Term and MARC 655 Index Term-Genre/Form.

The findings of this in-depth content analysis suggest several recommendations to the library metadata community, including creators and editors of RDA-based MARC 21 records, LC's Linked Data Service, OCLC, and developers of controlled vocabularies. Implementation of these recommendations will result in a stronger support of Linked Data in bibliographic metadata. The first recommendation is to include subfield $0 with an authority record ID number for all instances of field 043, which contains terms from Geographic Area Code controlled vocabulary (currently available through LC's Linked Data Service portal), field 655, which contains LCGFT genre headings (also available through LC's Linked Data Service portal), and field 648, which contains FAST chronological facet terms.[43]

Second, the authors advocate for revising the algorithm for automatic generation of FAST headings from those available in LCSH subject strings in fields 650 and 651 to also include fields 600, 610, and 611, which have chronological terms in subfields $d and/or geographical terms in subfield $c. This will result in automatically generating field 648 with chronological facet terms from FAST controlled vocabulary and field 651 with geographic name facet term from FAST controlled vocabulary. Another recommendation is to consistently include field 084 or field 072 with BISAC subject codes from BISAC controlled vocabulary whenever BISAC subject headings are used in field 650.

Finally, the authors recommend adding the most frequently used non-LCSH-based lists of subject headings—BISAC and SEARS—to the LC Linked Data Service Portal in Linked Data form with unique record IDs. When these controlled vocabularies become available in Linked Data form, it would be possible to add subfields $0 in field 650 instances that contain SEARS and BISAC subject headings.

The recommended steps are based on issues observed during in-depth analysis of records in this study, and addressing these deficiencies is expected to substantially improve subject access in general and Linked Data functionality of subject representation in bibliographic metadata in particular. The authors of this paper realize that in practical terms, the recommended steps will increase the workload of the cataloging agencies and would require additional resources to implement. While the authors believe (and many of their colleagues would agree) that the projected gains in subject access and Linked Data functionality support for the users are worth additional efforts, discussion is needed on the most logistically sound and cost-effective ways to approach these tasks.

Conclusion

This exploratory study aimed to address existing gaps in research and practice related to subject representation and Linked Data support in bibliographic metadata. It used in-depth content analysis of widely held RDA-based MARC 21 records in the WorldCat database. The study provides insight into the patterns of application of subject representation in the MARC 21 records created by libraries worldwide using the latest revisions of RDA and MARC 21 metadata elements to facilitate increased Linked Data functionality of library metadata. The findings indicate that the available MARC 21 content designation intended to support this functionality is not currently used to full capacity, and specific practical recommendations for addressing this gap are provided in the Discussion section of this paper. However, despite the observed limitations, overall, results of this study demonstrate that subject representation has substantially increased in extent and variety. The questions still remain about the extent to which this increased subject representation and Linked Data functionality supports the evolving user needs and impacts access to information. Future user studies will need to explore these questions.

The content analysis study presented here has several limitations that need to be addressed in future research. It is worth noting that the records analyzed were created in the first half of 2020 and therefore might reflect the changes in cataloging practices brought to light by realities of the COVID-19 pandemic and resulting adjustments to cataloging workflows to accommodate remote work. It is possible that because many catalogers in 2020 (as well as in much of 2021) worked remotely without direct access to non-digital collections, the emphasis shifted to refining existing records as opposed to creating new ones. A study of records created in the previous years but revised in 2020–21 might shed light on these trends and their effect on the completeness and overall quality of bibliographic records, and on

the degree of their support for Linked Data functionality. Another possibility is that the COVID-19-related adjustments in workflows resulted in scaled back cataloging with the intent to revisit it once catalogers were back on-site. Examination of the records created after the world largely emerges from the pandemic and those created earlier but last updated in late 2022 and beyond would allow to assess the impact of those trends on record creation and enrichment activity levels.

This study relied on a purposive sample of the one hundred most widely held (with five hundred or more holding institutions) RDA-based MARC 21 bibliographic records created in 2020 with the highest level of completeness. The study demonstrated that this group of records did not include records for materials in languages other than English or records that were cataloged in non-English languages. This limitation did not enable the authors to comparatively evaluate subject metadata, including application of Linked-Data-enabling subfields, for different groups of records based on the language of cataloging or language of materials. Future studies will address this limitation by analyzing large diverse samples of records. Additionally, there is a need to monitor the trends in subject representation practices, so future studies will compare the records created in 2020 with the records created in 2021 and beyond, including those created with the latest revision of RDA(3R). Moreover, comparing records analyzed in this study with the revised versions of the same records will help to trace changes in subject representation and, ideally, find more and stronger subject representation overall in our library information systems.

References

1. Ray R. Larson, "Between Scylla and Charybdis: Subject Searching in Online Catalogs," *Advances in Librarianship* 15, no. 1 (1991): 175–36.

2. Marcia J. Bates, "After the Dot-bomb: Getting Web Information Retrieval Right this Time," *First Monday* 7, no. 7 (2002): 1–2, https://firstmonday.org/ojs/index.php/fm/article/view/971/892.

3. Patrick Wilson, *Two Kinds of Power: An Essay on Bibliographic Control* (Berkeley: University of California Press, 1968), 69.

4. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, "Chapter 11: Subject Access," in *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015): 521–34.

5. RDA Steering Committee, *Resource Description and Access* (Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals (CILIP), 2010.

6. Library of Congress, *Bibliographic Framework Initiative*, (2021), accessed March 26, 2021, https://www.loc.gov/bibframe/.

7. RDA Toolkit, "Completion of 3R Project," October 15, 2019, https://www.rdatoolkit.org/node/202; Melanie Wacker, "Revised RDA Implementation Timeline: PCC Implementation Will Not Begin before July 2022," Program for Cooperative Cataloging List (PCCLIST), November 10, 2020, https://listserv.loc.gov/cgi-bin/wa?A2=ind2011&L=PCCLIST&P=10930.

8. *The Digital Future of the United States, Part I, the Future of the World Wide Web*, Before the U.S. House of Representatives, Committee on Energy and Commerce, Subcommittee on Telecommunications and the Internet, 110th Cong. (2007) (statement of Sir Tim Berners-Lee, CSAIL Decentralized Information Group, Massachusetts Institute of Technology), accessed March 25, 2021, https://www.govinfo.gov/content/pkg/CHRG-110hhrg39604/html/CHRG-110hhrg39604.htm.

9. Magda El-Sherbini, "RDA Implementation and the Emergence of BIBFRAME," *JLIS.it* 9, no. 1 (2018), https://www.jlis.it/article/view/66-82.

10. Diane L. Boehr and Barbara Bushman, "Preparing for the Future: National Library of Medicine's Project to add MESHRDF URIs to its Bibliographic and Authority Records," *Cataloging & Classification Quarterly* 56, nos. 2-3 (2018): 262–72; Carol Jean Godby and Ray Denenberg, *Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC* (Dublin, Ohio: Library of Congress and OCLC Research, 2015), https://www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015-a4.pdf; Jungran Park and Andrew Brenza, "Evaluation of Semi-Automatic Metadata Generation Tools: A Survey of the Current State of the Art," *Information Technology & Libraries* 34, no. 3 (2015): 22–42, http://ejournals.bc.edu/ojs/index.php/ital/article/view/5889; Jackie Shieh and Terry Reese, "The Importance of Identifiers in the New Web Environment and Using the Uniform Resource Identifier (URI) in Subfield Zero ($0): A Small Step that is Actually a Big Step," *Journal of Library Metadata* 15, nos. 3–4 (2015): 208–26; Jackie Shieh, "PCC's Work on URIs in MARC," *Cataloging & Classification Quarterly* 58, nos. 3–4 (2020): 418–27.

11. Jacquie Samples and Ian Bigelow, "MARC to BIBFRAME: Converting the PCC to Linked Data," *Cataloging & Classification Quarterly* 58, nos. 3–4 (2020): 403–17; Hyoungjoo Park and Margaret Kipp, "Library Linked Data Models:

Library Data in the Semantic Web," *Cataloging & Classification Quarterly* 57, no. 5 (2019): 261–77; Lihong Zhu, "The Future of Authority Control: Issues and Trends in the Linked Data Environment," *Journal of Library Metadata* 19, nos. 3–4 (2019): 215–38.

12. Marcia Lei Zeng and Phillipp Mayr, "Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-Dimensional Review," *International Journal on Digital Libraries* 20, no. 3 (2019): 209–30.

13. National Information Standards Organization, *ISO 25964 –the International Standard for Thesauri and Interoperability with Other Vocabularies*, 1.4, (2011/2013), https://www.niso.org/schemas/iso25964.

14. Library of Congress, *Subject Schemes* (2021), https://id.loc.gov/vocabulary/subjectSchemes.html. See the list maintained by the LC. The list currently consists of 371 items.

15. Book Industry Study Group, "BISAC Subject Codes FAQ," accessed March 25, 2021, https://bisg.org/page/BISACFaQ.

16. The Editorial Committee of the Chinese Library Classification, "Overview of Chinese Classification Subject Thesaurus," accessed March 25, 2021, http://clc.nlc.cn/ztfzfbgk.jsp.

17. Tina Gross and Arlene G. Taylor, "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results," *College & Research Libraries* 66, no. 3 (2005): 212–30; Tina Gross, Arlene G. Taylor, and Daniel N. Joudrey, "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching," *Cataloging & Classification Quarterly* 53, no. 1 (2015): 1–39.

18. Karen Smith-Yoshimura et al., *Implications of MARC Tag Usage on Library Metadata Practices*, a report produced by OCLC Research in support of the RLG Partnership (2010), 13, https://www.oclc.org/content/dam/research/publications/library/2010/2010-06.pdf.

19. Park and Brenza, "Evaluation of Semi-Automatic Metadata Generation Tools," 22–42.

20. Annif, "How to use Annif," accessed March 26, 2021, http://annif.org/.

21. OCLC, "Inside WorldCat," accessed July 12, 2021, https://www.oclc.org/en/worldcat/inside-worldcat.html.

22. OCLC, "Inside WorldCat."

23. William E. Moen et al., "Learning from Artifacts: Metadata Utilization Analysis" in *Proceedings of the Joint Conference on Digital Libraries–Opening Information Horizons*, (Chapel Hill, NC, June 11–15, 2006), 270–71.

24. Amy P. Eklund et al., "Comparison of MARC Content Designation Utilization in OCLC WorldCat Records with National, Core, and Minimal Level Record Standards," *Journal of Library Metadata* 9, nos. 1–2 (2009): 36–64.

25. Matthew Mayernik, "The Distributions of MARC Fields in Bibliographic Records: A Power Law Analysis," *Library Resources & Technical Services*, 54, no. 1 (2009): 40–54.

26. Smith-Yoshimura et al., *Implications of MARC*.

27. William E. Moen and Penelope Benardino, "Assessing Metadata Utilization: An Analysis of MARC Content Designation Use," in *Proceedings of the International Conference on Dublin Core and Metadata Applications: Supporting Communities of Discourse and Practice— Metadata Research & Applications* (Seattle, WA, Sept. 28–Oct. 2, 2003), https://dcpapers.dublincore.org/pubs/article/view/745.

28. Arlene G. Taylor and Charles W. Simpson, "Accuracy of LC Copy: A Comparison Between Copy that Began as CIP and Other LC Cataloging," *Library Resources & Technical Services* 30, no. 4 (1986): 375–87.

29. Shelia S. Intner, "Much Ado About Nothing: OCLC and RLIN Cataloging Quality," *Library Journal* 114, no. 2 (1989): 38–40.

30. Larson, "Between Scylla and Charybdis."

31. Herbert H. Hoffman, "Evaluation of Three Record Types for Component Works in Analytic Online Catalogs," *Library Resources & Technical Services* 42, no. 4 (1998): 292–303; Herbert H. Hoffman, "Subject Access to Works in Online Catalogs," *Technicalities*, 21, no. 5 (2001): 9–11.

32. Oksana L. Zavalina, Shadi Shakeri, and Priya Kizhakkethil, "Editing of Library Metadata Records and its Effect on Subject Access: An Empirical Investigation," in *Proceedings of the International Federation of Library Associations World Library and Information Congress Satellite Conference: Subject Access: Unlimited Opportunities* (Columbus, OH, August 11–12, 2016), https://pdfs.semanticscholar.org/4989/d3c7ba96a173117264e8f3e1310d0f4dcf0f.pdf.

33. Vyacheslav I. Zavalin, Oksana L. Zavalina, and Rachel Safa, "Patterns of Subject Metadata Change in MARC 21 Bibliographic Records for Video Recordings," in *Proceedings of the Association for Information Science and Technology* 58, no. 1 (2021): 1–5.

34. Vyacheslav I. Zavalin, "Analysis of Genre and Subject Representation in the Children's and Young Adults' Cataloging Program (CYAC) Bibliographic Metadata for Fiction Books," in *Proceedings of the International Association of School Libraries Conference* (Denton, TX, July 12–16, 2021).

35. OCLC Training and Support, "ELvL: Encoding Level," in *Bibliographic Formats and Standards*, accessed March 25, 2021, https://www.oclc.org/bibformats/en/fixedfield/elvl.html.

36. OCLC Training and Support, "042 Authentication Code (NR)," in *Bibliographic Formats and Standards*, accessed March 25, 2021, https://www.oclc.org/bibformats/en/0xx/042.html.

37. OCLC Training and Support, "042 Authentication Code (NR)."

38. Library of Congress Network Development and MARC Standards Office, "MARC 21 Format for Bibliographic Format: Appendix G: Format Changes for Update No. 31

(December 2020)," accessed March 25, 2021, https://www.loc.gov/marc/bibliographic/bdapndxg.html.

39. Shieh and Reese, "The Importance of Identifiers in the New Web," 208–26.
40. Wilson, *Two Kinds of Power*, 69.
41. Eklund et al., "Comparison of MARC"; Mayernik, "The Distributions of MARC Fields in Bibliographic Records"; Moen and Bernadino, "Assessing Metadata Utilization"; Moen et al., "Learning from Artifacts"; Smith-Yoshimura et al., *Implications of MARC*.
42. Pat Riva, Patrick Le Boeuf, and Maja Žumer, *Library Reference Model: A Conceptual Model for Bibliographic Information* (2017): 9, https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.
43. Library of Congress, "ID.LOC.GOV—Linked Data Service," accessed March 26, 2021, https://id.loc.gov/.

Book Reviews

Elyssa M. Gould

The Library Liaison's Guide to Collection Management. By Alison M. Armstrong and Lisa Dinkle. Chicago: ALA Editions in collaboration with ALCTS Publishing, 2020. 102 p. $49.99 softcover (ISBN: 978-0-8389-4802-6).

With *The Library Liaison's Training Guide to Collection Management*, Armstrong and Dinkle have created a thorough and thoughtful guide for library liaisons. Since much of this kind of work is specific to local institutions—which the authors keep in mind throughout the book—creating such a guide is no small feat.

Developed from the authors' own institutional liaison guide (Radford University), *The Library Liaison's Training Guide* is divided into eleven chapters. Each chapter is structured with an introduction to the topic, followed by more in-depth information and case study examples, ending with a lesson overview and "Local Practice Questions," which encourage the reader to ask about policies in their own library regarding these topics.

The book works best when defining the fundamentals of liaisonship, information that is particularly suited for individuals new to such responsibilities, as well as collection development librarians (CDLs) who may be required to train new liaisons or continuously adjust and review their institution's policies. The third chapter, "Fund Management" (17) is a perfect example of *The Library Liaison's Training Guide*'s strengths, with its valuable breakdowns of the types of orders (firm, standing, etc.), acquisition models, and budgeting suggestions that many libraries implement. The chapter that immediately follows on ordering is also a highlight, which manages to explore many nuances of the acquisition process while being broad enough to encompass different libraries' ordering procedures. Helpful too is the advice on collection assessment and weeding, which balances the considerations for adding and removing titles from a collection with excellent case studies. Chapter 6, "Building Relationships" (41) provides outstanding advice for the various ways liaisons can reach out to faculty, with tips on email etiquette and in-person interactions.

Some other particularly insightful sections include the advice on book selection in the second chapter: "Be aware of bias in your selections . . . consider whose voices are not represented" (11). Similarly, in the same chapter, the section titled "Equity, Diversity, and Inclusion in the Collection" (12) provides a generous list of sources from which liaisons can build their collection. While the fifth chapter ("Collection Development Committee") was written with collection development meetings in mind, many would benefit from the brilliant bullet points under the heading "Hallmarks of a Good Meeting" (37–38), which could be placed on the walls of every conference room in the country. The "Local Practice Questions" from each chapter are also compiled at the end of the book for easy reference.

However, some of *The Library Liaison's Training Guide* assets can be viewed as its drawbacks. The lack of specificity might be frustrating for new library liaisons, especially if they are not provided with proper training at their institution. The book might best be suited for CDLs interested in updating or developing training procedures. There is also a section in the second chapter regarding libraries ordering textbooks, which feels to be more specific to the authors and their university than actual guidance on the matter. Still, in the same paragraph, the authors go on to advocate for open educational resources (OERs), an avenue for which more libraries and liaisons should be advocating.

The Library Liaison's Training Guide to Collection Management is a great resource for any librarian interested in developing or improving their current collection development procedures, CDLs wanting guidance on training new librarians, and for novice library liaisons interested in mastering the complex and multifaceted components of their responsibilities.—*Cory Aitchison (cory.aitchison@pepperdine.edu), Pepperdine University*

Introducing RDA: A Guide to the Basics After 3R. 2nd ed. Chris Oliver. Chicago: ALA Editions, 2021. 216 p. $46.99 softcover (ISBN 978-0-8389-1908-8)

Introducing RDA: A Guide to the Basics After 3R is an updated version of the earlier edition, which covered Resource Description and Access (RDA) before the RDA Restructure Redesign (3R) project. Like the earlier edition, it does not provide detailed instructions for how to create bibliographic descriptions with the standard,[1] and it also does not provide an in-depth analysis or critique of the standards. Instead, *Introducing RDA* is something like a guidebook for RDA and the Toolkit. It aims to provide context for the standard and presents the vocabulary and basic concepts that will make it possible users to take advantage of the new official RDA Toolkit. It focuses on presenting the views of the RDA governing bodies and provides many useful citations. The new edition was necessary as the 3R project fundamentally changed how users interact with and navigate RDA and the Toolkit. It also redefined which types of information and instructions are found within the standard, and which are adjacent to the standard, but still available in the Toolkit.

The new edition covers the history of RDA from Anglo-American Cataloging Rules, 2nd edition (AACR2) through the revision and redesign of the standard and the toolkit. About half of the book focuses on the context for RDA, including the relationship with AACR2, the original RDA, and other international models and principles for bibliographic data. It provides little explanation about the influence of data models, standards, and technology outside of library traditions. As Oliver states in the preface, "this edition includes an overview of the major developments since the publication of RDA in 2010, focusing especially on developments that occurred during the 3R project: the impact of the 3R project itself, the results of aligning RDA with IFLA's Library Reference Model (IFLA LRM), and the outcomes of internationalization" (vii). A useful addition would have been a thorough analysis and explanation of the influence of linked data technology and data modeling on the redesign.

The volume starts with a general overview, and then introduces RDA as a standard intended to serve the international community. This section covers its alignment with existing international data models and standards, its approach to translations and the RDA governance structure. Related to internationalization, the next chapter focuses on the data models that underlie RDA. The chapter introduces the International Federation of Library Associations and Institutions (IFLA) Functional Requirements for Bibliographic Records (FRBR) family of data models. It then discusses the reasons for the creation of the IFLA Library Reference Model (LRM) and that model's primary attributes. Importantly, Oliver explains how the evolution to LRM is reflected in the changes in RDA after the 3R. Finally, for the introductory context, Oliver demonstrates the evolution from AACR2 through original RDA, to post-3R RDA.

Demonstrations of the transition from AACR2 to RDA are included in the fourth chapter, which could be useful even for those well-acquainted with the history. Oliver compares the treatment of the same data under AACR2, original RDA and RDA after 3R. This serves as an interesting view of both how the language has changed over the course of the development of these standards and clarifying the similarities and differences in the actual treatment of the data. For those interested in learning the new vocabulary and seeing demonstrations of how to cite and talk about the content of RDA, this is a useful introduction and model.

The last three chapters provide more of a guide to the actual content of RDA. These include a general overview of a few new concepts in RDA, including entities like nomen, aggregates, and timespans. Next is a guide to navigating and using the new the Toolkit. This section provides useful definitions, and the different options for recording data. Additionally, this section provides explanations of policy statements and application profiles and their role in RDA. Unfortunately, this volume was published before the Library of Congress and Program for Cooperative Cataloging provided policy statements and application profiles in the Toolkit. Including those would have provided concrete examples of how policy statements and application profiles that are found in the Toolkit, but outside of RDA, are intended to work and how RDA continues to meet some of the primary functions of AACR2.

Like a guidebook, *Introducing RDA*, ideally is not used linearly. Some of the information is repeated in various chapters. Additionally, some topics are mentioned early in the text as if the reader should already understand them. These same topics are then fully explained later in the text. As a reader approaches a concept of which they have little understanding, it will be worthwhile to refer to the index to locate a more thorough discussion and explanation. *Introducing RDA* models the use of language and citations of the standard that will allow newcomers to the now official Toolkit to navigate it with confidence. It is a useful guide for those coming from the Anglo-American cataloging tradition seeking to acquaint themselves with post-3R RDA.—*Jeanette Norris, jeanette.norris@yale.edu, Yale University Library*

References

1. Chris Oliver. *Introducing RDA: A Guide to the Basics*. (Chicago: American Library Association, 2010).

Telling the Technical Services Story: Communicating Value. Edited by Kimberley A. Edwards and Tricia Mackenzie, Chicago: ALA Editions, 2021. 229 p. $64.99 softcover (ISBN 978-0-8389-4946-7)

Professional discussions about the value of library technical services have long been the subject of the library literature and conference presentations. Often not fully understood until experienced first-hand, technical services work can appear mysterious to our library and organizational colleagues, and conveying the impact of our work is even more of a challenge to explain to library patrons engaged in teaching and learning. Because of this disconnect, many technical services librarians see advocacy as part of their role. The business of technical services evolves constantly in response to changing acquisitions models, data and discovery requirements and capabilities, and technological innovations, and therefore, it is no surprise that at the heart of advocacy lies communication. *Telling the Technical Services Story: Communicating Value* is a compilation of interesting case studies, primarily from academic libraries, that underline the importance of communication in conveying the impact of library technical services. Editors Edwards and MacKenzie note in their introduction that technical services work has moved beyond the back room and remains a critical part of fulfilling the library's mission. Due to the positioning of technical services, a communication strategy is often needed and sharing these practical ways to carry out this strategy is reflected by the publication of this title.

The fourteen case studies contained in this volume provide practical advice and share specific tools used to support communication by technical services librarians within units and departments, across organizations, and out beyond into the broader community. In her book *Academic Library Metamorphosis and Regeneration*, organizational development librarian Marcy Simons points to the marketing strategy of "seven times, seven ways" as a rule of thumb to communicate change.[1] Taking this rule and applying it to conveying value in technical services work, understanding as many creative ideas and innovative approaches as possible becomes necessary to build a communication approach that can be adapted to fit the local need. The volume is logically organized by scope with case studies about internal department communication first, followed by two parts that introduce case studies involving interactions with library stakeholders and outlining those designed to connect with the greater library community.

In thinking about the local, departmental view first, strong communication with clarity is the key. Four institutions provide insights into their efforts to strengthen the foundation of their communication within their units with an eye toward seeing these efforts as a "laboratory for experimenting with communication" (51) as shared by the University of Iowa in the context of their collection relocation project. A variety of outcomes emerge from these efforts, such as having a cohesive, formalized reference resource like that of the Michigan State University's documentation repository; managing workloads using project management as described by San Diego State University; or building trust by holding retreats as suggested by the University of Illinois at Chicago. The examples in part 1 point to the importance of communication transparency, accessibility, efficiency, sharing, and stewardship through what the University of Iowa calls "intentional communication" (58). Each case study also highlights the need for careful conceptual planning, concrete organization of the communication technique to be used, and follow through on implementation of the chosen strategy. Many techniques and output systems are mentioned, including follow-up surveys, content mapping, infographics, and sharing information via web applications.

A common theme of collective expertise follows in part 2, which focuses on communication between different library departments as a means to establish shared expectations. Five case studies reflect how to interact and intersect with the complexities of library technical services work. Often confusion or misunderstanding create barriers that can be addressed through the approaches shared in this set of stories. Colorado University-Boulder outlines its use of a collaborative project management application to create consistent, transparent and efficient workflows requiring participation by multiple departments. A wonderful "E-resources troubleshooting chart" (116–17) is included from Marymount University as they share the story of how they trained public services staff to assist with resolving access issues for electronic resources. Starting small with a community of practice helped librarians at Ohio Northern University make personal connections that were supported by stronger operational and planning communication. The University of South Florida-Tampa developed a formal training program for those outside the metadata operation who utilize the catalog on a regular basis and also established a shared vocabulary to facilitate conversations about further work together. The University of Tennessee-Knoxville used surveys and focus groups to bridge the communication gap between technical services and subject librarians to create trust and build stronger bonds between the units. These stories about cross-program communication consistently point to engagement, enlightened and informed members, and a goal of comprehensive understanding at their foundation and they result in efficiency and better service for patrons.

In the most outward facing scope and set of stories shared in this volume, communication is often combined

with other approaches, such as data driven analysis, marketing principles, or outreach techniques. Part 3 addresses this perspective as illustrated by five case studies that represent the added political and professional weight that comes with selling value to university administrators, faculty, and students. The Colorado School of Mines takes advantage of collection data to reinforce their story, framing data in the context of the library, its peers and the industry at large. Use of data also plays a large part in the Loyola Marymount University's case study about collection deselection. Programmatic decision-making is possible by using data created and maintained by technical services, combined with data from other sources and systems. The concept of developing an "(over) communication plan" (202) points to the nuances of balancing perceptions and hard data in a project setting that includes librarians and other stakeholders. Harkening back to marketing approaches referenced by Simons, such strategies form the focus of the case study shared by librarians from the Space Telescope Science Institute (STScI). Service development is the focus of the final two case studies from Georgia Tech and the University of Rochester-River Campus, with the former focusing on the user research process and portfolio management and the latter on building a program of metadata services that can serve the campus research community.

The words of contributing authors Mezick and Gould ring true for this volume, when they write, "When the context of what we do and what others do is widely understood, successful outcomes are produced" (136). While the changing nature of technical services as addressed in this volume is similar in topic and structure to other recently published works, such as *Library Technical Services: Adapting to a Changing Environment*,[2] this volume has the feel of a "how-to" manual specifically for technical services communications. Its chapters are relatively short in length but full of practical advice. Examples of checklists, planning techniques, organizational tools, schedules, and surveys are included for quick adaptation and adoption. Due to the practical nature of the volume, any librarian who needs to communicate the value of individual or group work could benefit from considering these case studies. Bibliographies and notes are also included to point the reader to more examples in support of the process outlined within each chapter. Even though authorship is heavily slanted academic libraries, the techniques and systems mentioned are universal.—*Laura Sill (ljenny@nd.edu), University of Notre Dame*

References

1. Marcy Simons, *Academic Library Metamorphosis and Regeneration* (Lanham, MD: Rowman & Littlefield, 2017).
2. Stacey Marien, *Library Technical Services: Adapting to a Changing Environment* (West Lafayette, IN: Purdue University Press, 2020).
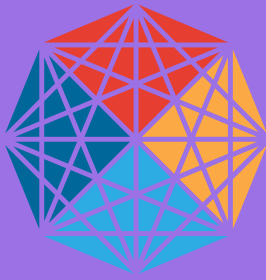
Core

LEADERSHIP
INFRASTRUCTURE
FUTURES

A DIVISON OF THE AMERICAN LIBRARY ASSOCIATION
225 North Michigan Avenue, Suite 1300, Chicago, IL 60601 ● 312-280-5038
Fax: 312-280-5033 ● Toll Free: 1-800-545-2433 ● http://www.ala.org/core/