4-2012

Cresting toward the Sea Change: Literature Review of Cataloging and Classification 2009-10

Sue Ann Gardner

University of Nebraska - Lincoln, sgardner2@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/libraryscience

Part of the Library and Information Science Commons


https://digitalcommons.unl.edu/libraryscience/265

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Cresting toward the Sea Change: Literature Review of Cataloging and Classification 2009-10

Sue Ann Gardner
Scholarly Communications Librarian, University of Nebraska-Lincoln, Lincoln, Nebraska, USA; sgardner2@unl.edu

Abstract

This review covers cataloging and classification literature published in 2009 and 2010, including relevant explorations of knowledge organization systems and theory. Only English-language literature is reviewed, though not all of the literature covered is U.S.-based. Overarching themes presented in the literature include the merging of library metadata into the Web environment, the continuation of cooperative cataloging in libraries, the role of both controlled and uncontrolled headings in catalog records, and reconsiderations of workflow in light of impending changes to cataloging rules. Notably, several relevant foundational documents were either completed or revised during the review period.

The literature covering cataloging and classification in 2009 and 2010 reflects the dramatic changes taking place in cataloging and shows that cataloging as a means to resource discovery is evolving. With nascent efforts to integrate bibliographic data into the Web environment, in the form of linked data and the Semantic Web, nothing less than a sea change is emerging. The literature reflects the many creative approaches being taken to adapt to this potential reality, such as experimentation with FRBR-ized catalogs, based on the Functional Requirements for Bibliographic Records (FRBR). The limitations inherent in the more than forty year old Machine Readable Cataloging (MARC) format have been identified and shared, paving the way for acceptance of a more interoperable format. Several foundational documents, such as the Statement of International Cataloging Principles and others, were either revised or completed during the review period, eliciting analysis and commentary. The use of controlled (i.e., subject, name, and series) headings and uncontrolled headings (i.e., tags) in records has been examined at length. Traditional concerns regarding cooperative cataloging and workflow also are well-represented in the literature, but often within the context of changes in the culture at large. Additionally, some excellent forays into the history of cataloging and classification were published.

Method

The author and an assistant identified some 450 possible publications to review that appeared in 2009 and 2010. Under the direction of the author, Anna Sophia Cotton, a recent graduate of the University of Missouri-Columbia School of Information Science and Learning, compiled citations from...
Cresting toward the Sea Change

August through October 2010 using Library, Information Science, and Technology Abstracts, and Google. Terms searched included: cataloging, bibliographical control, information organization, AACR2, RDA, MARC, authority control, classification, DDC, subject heading, LCSH, FRBR, metadata, OCLC, WorldCat, and Semantic Web. The author scanned the tables of contents for 2009-2010 issues of highly relevant journals to ensure that especially pertinent sources were revealed. The author subsequently performed circle searches by scanning the bibliographical references found in sources. The author also scanned tables of contents for appropriate publications listed in the Directory of Open Access Journals (DOAJ) under Social Sciences—Library and Information Science, such as Ariadne and Code4Lib Journal.

The focus of the selected sources is on the practice and theory of bibliographical cataloging and classification, including knowledge organization systems and theory. The author limited the universe considered to English language sources published as journal articles, conference papers, monographs, reports, and policy documents, electing those of greatest significant or representative of a topic. The volume of publications precluded being exhaustive. The literature review is organized into the following categories:

- Bibliographic Standards, Principles, Formats, and Codes
- Library Data in the Web World
- Workflow
- Cooperative Cataloging
- Personnel and Education
- History of Cataloging and Classification
- Classification
- Controlled and Uncontrolled Headings
- Cataloging Special Formats

Bibliographic Standards, Principles, Formats, and Codes

An internationally-constructed foundational document, built “on the great cataloguing traditions of the world,” the Statement of International Cataloguing Principles (ICP), was released by the International Federation of Library Associations and Institutions (IFLA). This concise “statement replaces and broadens the scope of the Paris Principles from just textual works to all types of materials and from just the choice and form of entry to all aspects of bibliographic and authority data used in library catalogues.” The ICP incorporates the entire framework of FRBR, includes discussion of the entities from each FRBR Group, and reiterates the functions of the catalog as outlined in FRBR, namely, find, identify, select, and obtain. Pointedly, the International Standard Bibliographic Description is referred to in the ICP as “the internationally agreed standard” for catalog record construction and display. Throughout the ICP, the convenience of the user is invoked.

Also fully incorporating the concepts put forth in FRBR was the draft consolidated edition of the International Standard Bibliographic Description (ISBD). This draft states that “the main goal of the ISBD is, and has been since the beginning, to provide consistency when sharing bibliographic information.” A significant addition to the 2010 consolidated edition draft is Area 0: Content Form and Media Type Area, which takes the place of the general material designation (GMD) formerly included in ISBD Area 1. The recently-created MARC21 fields 336 (content type), 337 (media type), and 338 (carrier type) will contain the information that resides in the new ISBD Area 0.

Several papers analyzed these two publications. Guerrini wrote two. One described in detail the new ICP. The other, co-authored with Bianchini, considers the ICP in relation to FRBR, the ISBD, and Resource Description and Access (RDA). Bianchini and Guerrini make the point that the ICP, the ISBD, FRBR, and RDA have been formulated by disparate bodies, and consultation among them would have produced a more unified suite of documents. They see the ICP, the ISBD, and FRBR as coextensive and resting on long-held theoretical bases for the organization of information. While they declare, “we should support the RDA effort,” they have many critiques of RDA, both in its construction and in its fundamental underpinning. In their most forceful plea, they believe that the ISBD should be acknowledged directly and prominently in RDA as the internationally-agreed upon standard, not merely for purposes of consistent display, but also for the “instructions for data analysis, that is, it gives stipulations to the cataloger to search for and recognize data, to define the functions of each data element within the specific context and to ascertain the proper position for recording the data element within the areas of the description.” They see the ISBD as more than a prescription for punctuation and placement of elements in a catalog record—it is the grammar of cataloging, giving catalog data meaning and, therefore, is essential to bring coherence to any cataloging standard.

Creider compares the Paris Principles with the 2009 ICP. He chronicles some of the political machinations that led to the ICP and extols the international nature of its development. He notes, “in sum, eighty-one countries participated in the various IME ICC [IFLA Meeting of Experts on an International Cataloguing Code] sessions,” and affirms, “this is great progress.”

Functional Requirements

Each of the three functional requirements documents was issued (at least in draft) by 2010. After eighteen years of development, a study group under the auspices of IFLA produced an amended and corrected version of FRBR in February 2009. This clearly-written report includes the background of the initiative to produce the document as well as a thorough description of all of the elements that comprise FRBR, namely, entities, attributes, relationships, and an explication of user tasks (which mirror the functions of the catalog,
as outlined above, namely, to *find*, *identify*, *select*, and *obtain*, and *basic requirements for national bibliographic records*.

Functional Requirements for Authority Data (FRAD) also was presented in a final form in 2009. FRAD is intended to be an adjunct to FRBR, further exploring and extending the Group 2 entities of *person* and *corporate body*. One main contribution of FRAD to the FRBR model is the addition of the element *family* to the Group 2 entities.

Functional Requirements of Subject Authority Data (FRSAD), released in June 2010, is highly conceptual and ultimately does not add much to the possible application of the general FRBR model. It suggests that the original Group 3 entities defined in FRBR, namely, *concept*, *object*, *event*, and *place* should remain, with no others added to the third group, though the authors suggest that Group 1 and Group 2 entities also be considered as potential subject entities. A key concept posited in FRSAD regards the definition of the phenomenon of *thema*, which are subject concepts, and associated *nomens*, which are signifiers by which the *themas* are known.

As an entity-relationship model, FRBR lends itself to Semantic Web applications. As Coyle says in RDA *Vocabularies for a Twenty-First Century Data Environment*:

The great value of using entities and relationships is that they allow the creation of a network of connections that goes beyond the description of a single item, more accurately reflecting the rich interaction between the intellectual creations that are being cataloged...it is these relationships that could transform library data into a true information network rather than a mere list of individual items.

Coyle acknowledges that the application of FRBR is in what she terms beta mode, but she is confident that it holds a key to the Semantic Web.

Explanation and analyses of the FRBR family of concepts abound and some testing has been done. Copeland gives a clear summary of FRBR and RDA and Gemberling examines the concept of *thema* and FRBR Group 3. Gemberling talks about FRSAD in a concise and straightforward manner, and addresses the oft- vexing problem of whether names for buildings should be established as corporate entities or as thematic subject entities. He shows how decisions regarding something such as this are arrived at arbitrarily and then the consequence of that arbitrary judgment gets carried forward with zeal, undeserving of the original intent behind the decision. His is a cautionary tale that should inform the eventual general application of the concepts outlined in FRBR, FRAD, and FRSAD.

**Machine-Readable Cataloging (MARC)**

MARC, as a legacy dataset, may never be fully extended into the wider Web, though some enterprising coders do their best to manipulate “a data format built to contain catalog records,” as opposed to bibliographic data in and of itself, as Thomale points out. Thomale, from a coder-turned-cataloger’s perspective, sees that MARC catalog records are structured data rather than data records. That is, MARC records replicate information on a catalog card, in that format, and do not neatly encode for discrete bibliographic elements, which would make them eminently more extensible in the Web environment.

The RLG Partnership MARC Tag Usage Working Group considered some aspects of the implications of MARC tag usage on library metadata practices. They discovered that of nearly 200 available MARC tags, 102 are used with significant frequency and eighty-six are little used or not used at all. Through their analysis and from user studies, they have determined that the following tags are the most meaningful to users or are very highly used (or both): 245 (title and statement of responsibility), 260 (imprint), 300 (extent), 1XX/7XX (main/addition entry), 65X (subject), 505 (contents note), 520 (summary note), 856 (online location and access), and 020 (ISBN). One conclusion they draw is that, “With more text indexed by search engines, focus should be on the authorized names, classifications, and controlled vocabularies that key word searching of full-text will not provide.”

Eklund and colleagues also conducted an empirical study of MARC content designation use in WorldCat records. Their goal was to determine if MARC tag usage mirrored required elements in national-, core-, and mini-mal-level records, and found that they do not. The authors recommend that empirical studies, such as theirs, be consulted when standards are formulated, to ensure that cataloging practices emphasize the useful aspects of a catalog.

Mayernik enacted a power law analysis of the distributions of MARC fields in the Library of Congress (LC) online catalog. With respect to bibliographic records, a power law analysis will show that a few fields will be used most frequently in records and the majority of fields will be used less frequently. Mayernik’s data show that the 1XX/7XX (main/addition entry), 245 (title and statement of responsibility), 260 (imprint), 300 (extent), and 6XX (subject) fields were used most frequently, and the 5XX (notes) and 4XX (series) fields were used next most frequently in his 1,500-record sample. He found that 76X-78X, 130, 240, and 250 fields were used infrequently.

Godby explains the specifics of crosswalking MARC with ONIX, EDItEUR’s Online INformation eXchange communication standard. She notes that “The two standards are structurally and semantically different because they support different needs and communities of practice.” Although because “the ONIX standard was proposed some thirty years after MARC was first adopted by libraries, it was informed by lessons learned from users of MARC.” As outlined in the Godby report, the set of semantic correspondences devised by OCLC staffers Renee Register and Bob Pearson is affected in part by the library community’s adherence to ISBD presentation. Register and Pearson discovered that MARC-to-ONIX and ONIX-to-MARC converse mappings can suffer due
Mayernik includes an important section in his paper on how RDA’s “organizational scheme draws on the conceptual models found in the FRBR and Functional Requirements for Authority Data (FRAD) reports. . . . The material-specific rules—such as those for music, recorded sound, video, etc.—are mixed into each chapter of RDA rather than the AACR2 practice of giving them their own chapters.”

To catalog an item of printed music, he notes, a cataloger would have to consult pages 25, 56, and 194 of RDA chapter 2, and pages 30 and 134 of chapter 3, whereas in the Anglo-American Cataloguing Rules, 2nd ed., (AACR2), all of these rules are found within the twenty pages of chapter 5.

Knowlton examines how legacy formats such as reproductions, facsimiles, and microforms will be cataloged using RDA. He has an excellent section on cataloging reproductions using earlier codes and outlines the multiple versions problem succinctly. He details the technical aspects of microforms that can be included in an RDA record, which are more extensive than those prescribed by AACR2 in many cases. Knowlton, like Mayernik, decries the “dispersal of applicable rules throughout the code. . . . [thus] a cataloger needs to consult up to six separate chapters and appendices to be sure of applying the correct rules to a reproduction in hand. There is no rubric, such as that found in each chapter of AACR2, to guide catalogers through general rules as they apply to particular formats.”

Hillman and colleagues wrote a paper containing details of the development and registration (in the National Science Digital Library Registry) of the RDA vocabularies that will likely be used in Semantic Web applications. They explain that “a key aspect of a registry is that it can provide a unique identifier (URI) for each data element and for each member of a vocabulary as well as one for the vocabulary or element set as a whole. With registered elements and vocabularies, labels can be identified for different languages or different communities, though the identifier can remain the same.”

Library Data in the Web World

Library metadata are currently highly textual in nature, as they are in the MARC format. MARC fields are not constructed like other machine-manipulable data, in strings of discrete data elements, arranged employing a simple syntax. Each MARC field employees varying syntax and cannot be machine-parsed consistently using algorithms. Library metadata will need to be more truly machine readable to be more readily incorporated into the Web.

If a reader is new to the concept of non-MARC metadata and the Semantic Web, a good place to start is Dunsire’s “The Semantic Web and Expert Metadata: Pull Apart Then Bring Together.” He includes a historical explanation of bibliographic metadata representation, from catalog card through to a Semantic Web schematic. He notes that “information professionals trying to create structured, accurate and comprehensive metadata cannot keep up using ‘traditional’ methods. Instead . . . we need to get our machines to process metadata as effectively as they process data.” Dunsire explains how triplets in the Resource Description Framework (RDF) are constructed, these triplets serving as the basic units of the Semantic Web, and he elucidates the components and mechanisms that are involved in Semantic Web development. He uses dozens of terms and acronyms in his paper and succinctly describes each one.

A good place to progress from Dunsire’s Semantic Web paper is to Coyle’s series of papers that were published in Library Technology Reports. Five of the six chapters in the January–March 2010 Reports provide a detailed overview of Semantic Web concepts, starting with the question “how can the library catalog move from being ‘on the Web’ to ‘of the Web’?”. Coyle answers this question, in part, in “Changing the Nature of Library Data,” by noting that catalog records need to be less textual and constructed of pure strings of data so that computers can recognize the data elements in them so those elements can be enfolded readily into the Web environment.

Yee offers a thoughtful exploration of whether bibliographic data can be put directly on the Web. She wonders about issues such as authoritative provenance. For example, are the data we will be using going to be vetted some way, and assuredly correct? She brings up many other concerns. Will data points be brought together to correctly describe a bibliographic entity? Will the human-machine partnership be adequately considered when constructing Semantic Web related protocols? How granular will the display of data be to the user? Yee points out that the concepts of entity/class and attribute/property are difficult to parse ideologically and will certainly, therefore, be difficult to implement in practice. She also points out that “bibliographic data [are] rife with hierarchy” and that she does not see how this issue is accounted for in Semantic Web development.

Tolkoff also lists some potential pitfalls with regard to interoperability including problems with transliteration and translation, as well as issues related to use of diverse subject heading systems internationally.
as the Virtual International Authority File (VIAF) and Multilingual Access to Subjects (MACS), if successful, could ameliorate some of these problems. Tolkoff believes that linking of these files could offer an elegant solution, though the currently-available interface is inadequate and will need improvement to be most effective.

After reading numerous papers on non-MARC metadata, the Semantic Web, and interoperability, a reader may be ready to take in the graphic depiction of the metadata universe, Seeing Standards: A Visualization of the Metadata Universe, that Riley and Becker constructed. Accompanied by a glossary of terms, the graphic groups 105 currently-used, cultural heritage-related metadata phenomena based on domain, function, purpose, and community. Riley and Becker’s Seeing Standards domains include: cultural objects, datasets, geospatial data, moving images, musical materials, scholarly texts, and visual resources; functions include: conceptual model, content standard, controlled vocabulary, framework/technology, markup language, record format, and structure standard; purposes shown are: data, descriptive metadata, metadata wrappers, preservation metadata, rights metadata, structural metadata, and technical metadata; communities listed are: archives, information industry, libraries, and museums. The graphic is valuable for its breadth and for its attempt at categorization.

To what extent has the cataloging community embraced non-MARC metadata creation and interoperability? As the Yang, Lee, and Xu article reveals, not much. Similarly, Park and Tosaka conducted a survey that shows that only 9.9 percent of the 263 total respondents used more than three schemata during the metadata-creation process. Ultimately, their results show that MARC continues to be the metadata schema of choice for digital collections.

Workflow

Papers by Ma, Toy-Smith, and Veve and Feltner-Reichert covered incorporating non-MARC metadata creation into catalogers’ workflow: “What We’ve Learned from the RLG Partners Metadata Creation Workflows Survey” is an assessment compiled from a 2008 survey of 121 Research Libraries Group (RLG)-affiliated librarians showing that “the tools being used are very localized, and no one tool kit is being used,” and “institutional routines are not yet standard enough for inter-institutional collaboration.” Ayers and her colleagues challenge the assumption that users will search using library sites instead of Web search engines, though the team learned that library data are available via multiple pathways, such as through various Web crawlers and the Open Access Initiative-Protocol for Metadata Harvesting (OAI-PMH) harvesters. Respondents reported that they are seeking to streamline workflows, especially for application of descriptive elements and for creation of procedures. The team concludes that this is still the initiation period for metadata creation in libraries and poses questions for further inquiry.

Valentino describes a project to integrate digital library metadata creation into the cataloging department at the University of Oklahoma Libraries. She details the process, including the training period and the testing period, which may serve as a template for others wishing to do the same. She concludes that cataloging staff are logical partners with digital library staff and can free the digital projects librarians to work more on digitizing and less on time-consuming metadata creation.

In “Mountains to Molehills: The Past, Present, and Future of Cataloging Backlogs,” Howarth, Moore, and Sze conducted a review of the literature on cataloging backlogs. One phenomenon they mention is something that Sarah Thom called “bibliographic chicken,” in which catalogers leave an item in the backlog while they wait for someone else to create an original record for it and contribute the record to a shared database. The authors point out that not all library managers wish to eliminate backlogs entirely because they offer a reserve of materials to work on when acquisitions have slowed.

Co-commissioned by the National Information Standards Organization (NISO) and OCLC Online Computer Library Center, Luther prepared a well-written and informative white paper, Streamlining Book Metadata Workflow. This report encapsulates the methods used by publishers, metadata vendors, book sellers and wholesalers, national and other libraries, and Google, to exchange book-related data, primarily in ONIX and MARC. Luther includes statistics about the size of the market and information about how the stakeholders in the sector work together, though their metadata operations have not been fully standardized yet. For example, Nielsen Book, a commercial metadata vendor in the United Kingdom processed 43.2 million records in 2008, the vast majority of which involved updates on price and status (the status might generate a “Hurry, there are only 3 left!”-type of message that a potential purchaser would see on a product page online)—critical elements for commercial entities, that must be absolutely current at all times. Luther reports that “OCLC hosted a Symposium for Publishers and Librarians to explore metadata needs and practices,” including the issue of various identifiers (such as the ISNI (International Standard Name Identifier), ISBN (International Standard Book Number), DOI (Digital Object Identifier)), series, related works, and subject schemes. Many creative ideas for best practices were generated during the symposium.

WorldCat Local is the focus of a paper by Zhu of Washington State University Libraries, where they have integrated WorldCat WSU with their Innovative Interfaces integrated library system. Zhu lists the technical issues associated with configuring WorldCat Local and outlines the specific points at which technical services personnel were involved with the implementation of the product. Some local cataloging policies and procedures were reconsidered in light of the switch and the Griffin catalog had to be prepared in a variety of ways. Zhu feels that migrating to WorldCat Local will al-
low deferred projects to move forward, such as cataloging of hidden collections, retrospective conversion, digitization, and populating the institutional repository with content.

Kim describes an outsourcing effort at the Hanyang University Library in Seoul. The impetus for this project was to eliminate employees who conducted the “nonproductive and insignificant repetitive work” of cataloging and “to alleviate the bottleneck in employee promotion.” All library-employed catalogers were fired and replaced with contractually-hired employees. Library staff supervised the new cataloger, and the replacement catalogers occupied the same space in the library as the former employees. Kim says, “since the contract workers are not library employees, they are not highly interested in the affairs of the library,” and “since [they] work according to the conditions specified in the contract agreement, it would be difficult to expect sincere participation from them.” Kim says that “rather than forcing the library’s abstract vision onto the employees, the management should communicate with them to share its specific goals and vision,” then declares in turn that “every member of the department must share Hanyang University Library’s vision of becoming ‘the foundation for providing the latest knowledge and information for the nurturing of global leaders.’”

Cooperative Cataloging

In separate papers, Schuitema and Sellberg examine the likely role of cooperative cataloging in a post-online public access catalog (OPAC) world. Both take a historical view and Schuitema examines selected landmark developments. She wonders if librarians have reached an impasse in cooperative cataloging. Sellberg envisions librarians working cooperatively not to maintain and catalog redundant collections in each separate library but, instead, she says, “if one thinks about pooling the expertise of metadata experts and working together to facilitate use of the world’s information resources through well-designed and well-managed systems of access, then we have barely begun to realize the possibilities.”

Proving that cooperative cataloging activities are active in the United States today, articles by Banush, Charbonneau, and El-Sherbini cover the Program for Cooperative Cataloging (PCC) and its associated initiatives, BIBCO (Monographic Bibliographic Record Program), CONSER (Cooperative Online Serials Cataloging Program), NACO (Name Authority Cooperative Program), and SACO (Subject Authority Cooperative Program), and other cooperative cataloging ventures. Citing funding challenges in light of continual changes to information technology and the library community’s inability to adapt quickly enough, Banush cautions that PCC is in a vulnerable position. He believes that NACO and SACO, with their purpose devoted to authority creation, may hold promise as one viable future for the PCC, and stresses that “the Program will have to change . . . if it chooses to be an important influence on the future of bibliographic control.”

Spiteri used the heuristics of communication, identity, and perception in examining and evaluating the social features and completeness of the catalog records of sixteen social cataloging websites. She found that “although the bibliographic content of most of the catalog records examined was poor when assessed by professional cataloging practice, their social features can help make the library catalog a lively community of interest where people can share their reading interests with one another.”

The economic aspects of cataloging, as they relate to cooperative endeavors and other considerations, have come under scrutiny. The LC commissioning of R2 Consulting to conduct a study of the North American MARC records marketplace is evidence of this. Nine hundred seventy-two libraries and seventy vendors contributed to the R2 study data set. Fischer and Lugg reveal that LC records are underpriced, the market does not pay well for original cataloging, excessive editing of records for use in cataloging silos occurs, and librarians are unaware of restrictions on MARC record use or redistribution. The authors point out that “the prevalence of open databases is a key factor in the economic confusion that plagues the MARC Record Market.”

Fischer and Lugg distinguish between the community value system that drives library operations and the commercial values of vendors, and report a disconnect between the desired activities of the cataloging community and the true costs associated with those activities. They note that the LC is expected to perform a role as an uber-provider of cataloging output that it cannot continue with its projected budget. The report also emphasizes that the LC’s cataloging contributions are critically important to school and public libraries, which often have limited or no in-house cataloging expertise.

Regarding cooperative cataloging in the United States, a mere ten libraries contributed two-thirds of the BIBCO records in 2008. Fischer and Lugg point out that “a cooperative system only works well if everyone participates.” Additionally, according to their calculations of numbers of catalogers in North America, “we should not have backlogs.”

Wolven challenges every aspect of cataloging mores, among them the fiction that librarians catalog items based on what users want. He points out that librarians actually catalog what he terms units of commerce, such as books and serials, whereas patrons are really interested in specific poems in the books, certain articles in the serials, a particular song on a recording, and so on. Wolven suggests that “the scope of the library catalog that eventually emerged wasn’t shaped by user needs . . . but by technological and economic limitations,” and continues, “the incentive for libraries to trade descriptions of widely distributed published books drove standardization in ways that didn’t apply to competing publishers of indexes and bibliographies, or to manuscripts and other unique materials.” He frames his comments in light of the changes on the horizon for cataloging and believes that “gradually, we will probably reach a new consensus on best
practices, less grounded in 20th-century publishing patterns,” and pronounces that “we must recognize that the questions are changing.” 68

Personnel and Education

The evolution affecting bibliographic cataloging and classification extends to personnel. Those formerly called catalogers are increasingly referred to as metadata librarians and similar labels. This leads Hruska to consider, “Where Are We with the Staffing Transition from Cataloging to Metadata Management?” 69 She answers her own question in part when she says, “I believe that those operations that have traditionally been considered cataloging are well along in a transition that parallels the transformation of the content of library collections.” 70 She continues, “there is a marked shift in research libraries to focus more staff effort on including products of local digitization in an integrated discovery experience, moving beyond the silo of a library catalog. It is this work that is actualizing metadata construction as an essential part of library operations, describing the growing collections of digital objects.” 71

Mitchell, Thompson, and Wu describe their experiences with transitioning technical services staff to better reflect emerging models of discovery. 72 They say they want a more fluid dynamic among the personnel in technical services so that staff are able to take on new roles readily, with autonomy and with the proper training behind them. At the University of Northern Colorado (UNC), Leffler and Newberg report that several possible organizational structures were considered after a key staff member in technical services resigned unexpectedly. 73 With broad staff input, UNC selected a final organizational model based on research conducted to determine future directions in technical services there.

Three studies examined the changing role of the metadata librarian. Each employed the technique of quantifying elements in job descriptions to assess trends. In their paper, Han and Hswe explain that they wanted to examine the different responsibilities and competencies that metadata librarians have compared with those of cataloging librarians. 74 They looked at job descriptions that were in advertisements for open positions between 2000 and 2008 and discovered that the skill set required of metadata librarians has evolved over that time, indicating that more than a name change is occurring. The other two studies were conducted by Park and her colleagues. 75 In both, they looked at advertisements posted in the mid-2000s, used co-term and cocitation analyses to construct co-occurrence matrices, and visualized these findings through multidimensional scaling and cluster analyses. In addition to noting that managerial-level positions are in high demand, they discovered that “the advancement of technology has affected every aspect of the cataloging profession: job titles, competencies/skills, and responsibilities.” 76

Even for those whose titles remain simply cataloging librarian and variations thereof, changes are afoot in their daily work. Hitchens and Symons give a well-thought-out, detailed template for RDA training for catalogers at all levels. 77 They consider many pertinent issues and make suggestions such as mapping commonly used rules from AACR2 to analogous ones in RDA. They point out that experienced catalogers will need to get into the habit of looking up rules in RDA for things with which they were very familiar with when using AACR2.

Cox and Myers conducted a 237-response survey of staff at Association of Research Libraries member libraries showed that one central perceived difference between paraprofessional and professional catalogers was expectation for involvement in professional development activities. 78 Professional catalogers frequently report that they are expected to produce articles and other research output and to serve on committees, while almost no paraprofessionals report having such expectations placed on them.

Two articles reference paraprofessionals’ focus on training issues. Sapon-White gives a well-considered plan for training paraprofessionals to perform subject assignment of electronic theses and dissertations. 79 Shrinking budgets adversely affect numbers of professional catalogers in academic libraries, so paraprofessional staff are left to perform higher level work. He details the pitfalls that can arise and gives measured suggestions for addressing the various issues. Using the skills she acquired as a German-language teacher, Valente describes an approach to training paraprofessionals that focuses on the beginning days of employment. 80 She includes long lists of catalogers’ tasks and required abilities and includes some examples from work sheets that can be used in training.

Hudson considers the teaching of classification from 1990-2010 and is heartened to see that classification and knowledge organization are still taught in library schools, citing others who also feel it should remain in the curriculum. 81 Of the many issues that have influenced the teaching of classification, one that stands out particularly is the change brought about by the use of online tools, like Classification Web and WebDewey. Hudson believes a focus on the theory of classification is necessary in library schools so a student can make informed decisions when he or she is applying classification schemes.

Some articles concern pre-professional cataloging experiences. In a paper with a comprehensive literature review on the topic, a team from Kent State University (KSU) Libraries reports that the Libraries have a formal program for students of the KSU School of Library and Information Science. 82 The program is beneficial to both the Libraries and the students because, although the program has associated costs, the practicum students are not paid so the library sees a net cost savings. The students get formal training and hands-on experience cataloging before they complete their degrees. An apprentice cataloger gives his perspective at the end of the paper.
History of Cataloging and Classification

Some authors chronicled cataloging and classification history. Beall and Mitchell detail "the history of the representation of the Dewey Decimal Classification (DDC) in the . . . MARC formats, with special emphasis on the development of the MARC classification format." As of 2009, the DDC is finally represented fully in the MARC21 format. They explain that concomitant incorporation of DDC fields into the authority format was rejected, however, as the authority format is not extensible enough. With recent changes to MARC, DDC facets can now be shown. Work is underway to expose DDC in the Semantic Web, with uniform resource indicators (URI) scheme development and other explorations.

Mellwaine "traces the history of the Universal Decimal Classification (UDC), its relationship with other schemes, and opportunities for further collaboration." In summarizing the quest for a "universally acceptable system for retrieval of subjects," she avers that "the problem is to create a structure that is both universally acceptable and sufficiently detailed to be useful without being overly complicated." She explains that even the DDC, which is used worldwide, is not the ultimate system, and owes its success to external factors as much as to its robust construction. She discusses the early days of the UDC, whose development was interrupted by the First World War, and had a proviso placed on it that it would not—until much later—be published in English. She concurrently addresses the history of the DDC in her paper and describes how the two systems were developed, oftentimes at odds with one another.

Slavic contributed a chapter in Library and Information Science in the Digital Age: Essays in Honor of M.P. Satija in which she details the history of shelving and call number evolution for bibliographic collections in Europe. She explains that even something as mundane as labeling of books took much wrangling to devise. She notes that it was not until late in the nineteenth century that call numbers were based on systematic book arrangement determined by a classification scheme.

In "Wholly Visionary," Yee examines the history of the American Library Association (ALA) and the catalog card distribution program at the LC. She remarks, "the ALA is now dominated by library administrators with shrinking budgets who know very little about the complexities of bibliographic control (other than its expense) and who wonder if the fact that undergraduates are in love with Google might not provide an excuse for libraries to dispense with the information-organization part of their budget entirely." Yee draws on the context of past challenges and opportunities to show that the decisions made today about how much access to provide via cataloging practices will have a real, on-going affect on people. She encourages librarians to be visionary and bold in their mission to serve the public as a whole and not kowtow to commercial forces, nor bow to uninformed leadership.

Knowlton gives an overview of a debate that took place in the pages of Library Resources and Technical Services in the 1950s and 1960s, just before the Anglo-American Cataloguing Rules (AACR) were enacted in 1967. Seymour Lubetzky and others exchanged their thoughts on the best course of action over several years. Codes such as AACR do not arise out of a vacuum and with consensus. Knowlton notes that disagreement was resolved "in Paris, [where] representatives from thirty-four national library associations met and agreed on the Paris Principles, which served as the basis for future cataloging codes in most countries."

Elkington details some of the ground-breaking research conducted by OCLC, much of which is ongoing and affects many day-to-day cataloging activities. OCLC helped to spearhead the creation of Dublin Core (DC), when they convened a workshop in 1995 with the aim of envisioning a "simple, modular, extensible metadata scheme for Web-based resources." With various partners, OCLC also was a key player in the formation of PREMIS, the preservation metadata framework, VIAF, the Virtual International Authority File, and WorldCat Identities, a project that creates an automatically-generated summary page for every name in WorldCat, nearly 25 million names.

Classification

Classification in practice was the focus of a few articles. Mitchell and Vizine-Goetz chronicle the acquisition of the DDC by OCLC in 1988 and what OCLC has done to maintain it, including facilitating translations of it, and mappings of it with subject headings systems. Since the acquisition, OCLC researchers have conducted research on DDC-related topics, including attaching URIs to elements in the DDC, and modeling it in Simple Knowledge Organization System (SKOS).

Bibliographic classification on the Semantic Web was considered at the Dublin Core Conference in 2009, when Panzer and Zeng revealed some problems with modeling classification schemes in SKOS, particularly with so-called centered entries and number spans in the DDC hierarchy. They present the issues in detail in their paper and suggest possible work-arounds.

The DDC was the subject of a highly technical and comprehensive study undertaken by Wang, who explored the limitations and possibilities involved in the successful application of text categorization with the aim of automated assignment of DDC classes for bibliographic items. After much experimentation, Wang sees the DDC, above all, as a human construct that is based on convention rather than a machine-friendly hierarchy. Wang sees the need to rework the scheme to satisfy the needs of the computer in order to afford the DDC more interoperability. Classification theory was considered by some authors during the review period. Gnoli parallels, but more fluidly fleshes out, the ideas presented in the FRSAD document in his com-
mentary on the distinctions between phenomena and their attendant signifiers. Gnoli suggests that discrete phenomena encompassing a theme, for example, animals, ought to be taken to be the “basic units of classification,” because, “while disciplines are a traditional way of organizing knowledge, in many cases they also act as a superstructure adding unnecessary perspective to the content itself.” In the manner that classification schemes are ordered currently, he explains, animals, for example, are considered relative to society’s relationship with them at any point, such as whether they are pets, livestock, wildlife, fossils, or laboratory subjects. Gnoli suggests that animals be presented in a classification scheme simply using their chosen signifier (such as the English word animals and equivalent words in other languages) and that relationships are separately explicated in the classification scheme. The signifiers for the phenomena and, separately, the relationships could then be combined in endless ways, thus freeing the classification scheme from imposed perspectives.

Gnoli’s thesis ties in with what Lee reports in “Divination and the State: Classifying Technical Texts in Han China.” Lee quotes Tsuen-Hsuan Tsien to point out that “ancient writings in China were used for communication not only among human beings, but also between human beings and spirits.” This act was not akin to current-day spiritual explorations, but more like contemporary scientific investigations. Lee also explains that the issuing body in ancient China was often the most salient aspect of writings. Similar to the U.S. Superintendent of Documents (SuDocs) arrangement, whatever written products a certain governmental body issued would be grouped together by virtue of their authorship. Understanding the context in which ancient texts were constructed is paramount in classifying them meaningfully. Without a modern-day equivalent to the activities from earlier times, current classification schemes may offer no proper place for those ancient texts. Gnoli’s proposed construct is one possible solution to this problem.

Not only ancient texts but also those created more recently may not find a place in current classification schemes. The ever-increasing corpus of Islamic texts is one example. Idrees and Mahmood conducted a small but well-thought-out survey of ten Islamic and ten information science scholars asking how they believe an adequate Islamic classification might be devised because none exists. The DDC number for Islam, for example, is simply 297. The majority of the scholars in the study favor the idea of creation of an extensive and current classification scheme for Islamic literature.

Some public librarians in the United States are reconsidering application of classification schemes in their libraries’ book arrangements. As Fister reports, some have taken to modeling their book layout according to a version of the Book Industry Standards and Communications (BISAC) system. Fister quotes a blogging mom who declares “the books, everywhere, but especially in the children’s room, have been shelved, labeled, and organized in a way that makes me feel less like a moron and more empowered to find what I’m looking for on my own.” According to the survey conducted by Fister, many public librarians share this sentiment, at least to some extent, with more than 85 percent acknowledging that public librarians would better serve patrons if a not-strictly-Dewey system were implemented in their shelving schemes.

Controlled and Uncontrolled Headings

Some authors examined controlled headings in general. Hearn looked at library catalogs at universities that are part of the Committee on Institutional Cooperation (CIC) and the LC catalog to determine how much quality assurance of controlled headings is taking place in each. He found a wide variation in the pace of upkeep of changed headings between the catalogs, concluding that a study of this kind can help institutions in a consortium or other selected group of libraries to measure catalog quality based on comparison with that of peers.

In anticipation of adoption of RDA, Burke and Shorten chronicled practices before the cataloging standard changes and compare how authority work is being done in U.S. libraries today. They find that authority work varies more based on a library’s size than on the type of library. Larger libraries outsource their authority work more often than is done at smaller libraries and larger libraries are more often NACO contributors. They also found that personal name headings were the most frequently controlled and that uniform titles were least likely to be constructed or edited.

Names

“The DeathFlip Project: Automating Death Date Revisions to Name Headings in Bibliographic Records” is an undertaking of the librarians at Kent State University in their response to the LC’s decision to add death dates to name headings as they arise. In early 2006, the LC began to add death dates to name authority records, changing about 500 records per month the first two years after the rule revision. Kent State staff used the OCLC RSS (Real Simple Syndication) feed of the headings changes to initiate a protocol in their catalog to automatically flip the records so the proper, newer headings are incorporated into the catalog.

Dragon and her colleagues have made a concerted effort to control the name headings associated with the Eastern North Carolina Postcard Collection at East Carolina University. Under the premise that “minimization of the necessity of relying on chance for information discovery is the mark of quality metadata,” their thoughtful method includes many important points, so that any other group wishing to control name headings for a small collection of images would do well to reference this paper.
Subjects

Chung, Miksa, and Hastings explored the use of text categorization (TC) to assign subject terms, using algorithms that are modified with each successive use. They contend that humans do not use the full text when assigning subject headings, so computer algorithms perhaps should not, also. They argue that conceptual frameworks based on human-informed processes can best direct TC by computers. They discovered that keyword was shown to be a more effective source than full text when mining the article for subject terms, and cited that works, article title and journal title were as effective as mining the full text when assigning headings based on performing TC on a document.

Yi and Chan explored whether LC Subject Headings (LCSH) may be employed as an effective subject access tool in a networked environment. They examined the syntax and structure of the headings and determined that there is both a local relational structure as well as a global hierarchical structure. They conclude that the LCSH retains too many features of natural language to be used successfully as an interoperable controlled vocabulary and the LCSH needs to be more rigorously hierarchical. Ultimately, the syntactic structures are too diverse to parse algorithmically, making the LCSH of limited use in the networked environment.

Series

Sapon-White recounts how the cataloging staff at Oregon State University responded when the LC changed its policy to no longer trace series. They undertook the task of counting bibliographic records added over the year-and-a-half study period and noted the source of the records. The set of records with untraced series statements gathered from the LC during the study period was negligible (sixty out of 53,911). Only approximately 900 more than that came through with untraced series in total, thereby indicating that this change was not a burden on the staff and did not warrant a significant change in workflow.

Uncontrolled Headings

Uncontrolled headings are keywords attached to documents, websites, bibliographic records, or other content, that are not from a thesaurus or other list of controlled headings. The term “tag” is often used to denote keywords input by users of content, though expert metadata creators also can assign tags to a document. The aggregate of tags associated with a defined grouping of content is termed a folksonomy. Much research has been done to examine the use of uncontrolled headings in catalog records.

Bianco conducted a survey through which she found that social tagging has not been adopted widely by medical librarians. In another study reported by Maggio and colleagues, instructors asked medical students at Boston University’s Alumni Medical Library to assign tags to several digital objects. They asked the students in an Introduction to Biomedical Literature class to answer the question, “What would you call it?” The exercise elicited synonymous terms, spelling mistakes and variations, and variations in specificity, thereby highlighting the importance of using controlled terms when searching for medical literature.

Griffis and Ford demonstrated that subject liaisons can act as partners with catalogers to provide helpful keywords and descriptions of resources of electronic databases and media items. They observe that subject experts can offer high-quality user-generated uncontrolled terms in records that will enhance discovery. The article offers useful suggestions about embarking on an enterprise such as theirs.

Similarly, Strader examines “the overlap between author-assigned keywords and cataloger-assigned LCSH for a set of electronic theses and dissertations in Ohio State University’s online catalog.” She reports that the combination of controlled and uncontrolled terms in records improves retrieval, and notes that previous studies have consistently shown the same phenomenon.

Adler conducted a comparative study of controlled terms in LCSH and user-generated tags in LibraryThing for transgender books. She declares that “perhaps the greatest power of folksonomies, especially when set against controlled vocabularies like LCSH, lies in their capacity to empower user communities to name their own resources in their own terms.” Her research, comparing tags in LibraryThing for transgender-related books with LCSH assigned to the same books, revealed “a disconnect between the language used by people who own these books and the terms authorized by the LC and assigned by catalogers.” She concludes that the two sets of vocabulary each have their strengths and their limitations and thus complement each other.

Thomas, Caudle, and Schmitz also analyzed the tags assigned by users in LibraryThing. They found that “tag variations are the most prominent hindrance to search and retrieval,” variations referring specifically to, “tags which were the same except for tense, symbols, spelling, and capitalization . . . also included noun-adjective combinations and word combinations meaning the same thing or concept,” and acronyms and initialisms. Ten books were chosen for the study to which 7,653 tags had been assigned and 59 percent of those tags were either variations or contained non-alphabetic characters. The authors determine that folksonomies can augment controlled headings, but cannot replace them.
current ontology. She developed a model consisting of an ontology of a folksonomy in which relationships between tags, the objects they represent, and the tag assigners, are explic- ed. These relationships form patterns that then can be exploited and integrated into searching systems to better improve precision and recall.

Likewise, Kakali and Papatheodorou observe that “the current state of the art on the semantic correlation between folksonomies and knowledge organization systems (KOS) centers upon the analysis and integration of user, creator, expert, and machine generated vocabularies.” The authors direct- ed a tagging effort by a group of experts and had a group of catalogers inspect the tag choices. The authors also gathered tags from LibraryThing to supplement the tag set. The librarians approved of the majority of the tags input by expert users and a smaller quantity of the LibraryThing tags and found that tags supplement the subject heading assignment and improve searching.

Lawson also concludes that “while social tagging does consist of a great deal of subjective tagging, there is enough ob- jective tagging available on bibliographic-related web-sites such as Amazon and LibraryThing that librarians can use to provide enriched bibliographic records,” and recognizes that user-generated tags can enhance subject cataloging. Rolla concurs as he remarks in his paper on the topic, “user tags can enhance subject access to library materials, but they cannot entirely replace controlled vocabularies such as the LC subject headings.”

Cataloging Special Formats

Several helpful publications address the cataloging of specific material formats. Representative articles are reviewed.

Boock and Kunda compare past processes and work-flows for print theses and dissertations with the current workflow for those that are electronic. Part of the workflow involves student authors themselves entering basic metadata when they deposit their dissertations. Some of the metadata are generat- ed automatically via DSpace, and the library staff add subject headings and verify the student-submitted metadata. Helpful- ly, the authors include some cost-saving estimates that their workflow proffers.

Wu and Mitchell describe how difficult using vendor-supplied metadata for hundreds of thousands of e-books from several different vendors is. They use MarcEdit and the SerialsSolution MARC service for e-books to accomplish some cataloging tasks, some of which are batched, and they discuss the benefits of the provider-neutral record approach, namely, ultimately fewer and more-consistent records. They conclude by saying that e-book cataloging on the scale they describe is new for them, but that, with efficient solutions and an awareness of the resource supply chain, they have been able to absorb the increased throughput of materials. That said, they point to issues that still need to be resolved.

Recognizing the ongoing challenge to provide access to ever more digital resources Reese wrote “Automated Meta-
data Harvesting: Low-Barrier MARC Record Generation from OAI-PMH Repository Stores Using MarcEdit.” Reese points out that metadata arise from many sources, not just OCLC as was more common in the past. In describing two potential use cases, Reese shows how MarcEdit “offers default conversion support from OAI-PMH metadata to a num-

ber of different metadata formats.”

Beamer thoughtfully examines problem areas in map cata-

looging and, despite the inherent difficulty in cataloging maps, implores would-be map catalogers to “secure appropriate sys-
tems for retrieval and include geographical location informa-
tion, specifically numerical co-ordinates.” She looks not only at MARC21 and AACR2 as cataloging tools, but also Encoded Archival Description (EAD) and the Dublin Core in the Resource Description Framework format. She reports about the benefits and detriments of the proprietary Royal Scottish Geographical Society’s Images for All project’s Or-
acle-based system. Though imperfect, the United Kingdom Ministry of Defence Parsons Classification is used to denote locations in the system, which is useful because it includes non-extant geographic place names. Beamer also discusses the European Library’s open source DIGMAP retrieval system.

Kowal and Martyn consider cataloging a different set of maps, namely those found in books. This kind of granularity in cataloging becomes feasible when digitizing collections, where image-based information can be easily uncovered and highlighted. After the theft of pages of maps from books, the Vulnerable Collections Item Project was initiated at the British Library and more than 3,000 maps from the fif-teenth through the seventeenth century were selected. Maps were cataloged using MARC21 and AACR2 and the analytic records were linked to the parent record in each case. Using the ADAM module to link administrative and rights meta-
data to the catalog record, digital images of the maps also were linked to the records.

A group of Dutch and French librarians present “an exper-
iment on enhancing the semantic interoperability of two dig-
tal iconographic collections: Mandragore, the iconographic database of the Manuscript Department of the French Na-
tional Library (BnF), and the Medieval Illuminated manus-
cripts collection of the National Library of the Netherlands (KB).” They believe that the cultural importance of these collections and their similarity justify that “these collections need to be interconnected and made interoperable, in a way ‘smart’ enough to allow users to seamlessly interact with the resulting aggregates.” They report about an experiment they conducted that employed Simple Knowledge Organization System (SKOS) and an ontology matching technique that automatically identified semantic correspondences based on lexical alignment. They then built an interface to test the mod-
el and found that the results indicate that they could achieve their goal of interoperable searching using this method.
Conclusion

As 2010 was declared to be the Year of Cataloging Research by Carlyle and others, this was an especially auspicious time to review the literature. Despite the enormous volume of literature published during the review period, some themes merit highlighting. A common research topic was the encroachment on traditional cataloging by non-MARC metadata and the potential for bibliographic metadata’s interoperability with the wider Web, the end-product of which is often referred to as the Semantic Web. Another popular topic was the use of uncontrolled tags in catalogs, often user-generated, invariably showing that tags supplement, but do not supplant, the use of controlled headings in records. The literature reflects a continued emphasis on cooperative cataloging, what it has meant in the past, and what it will look like in the future. Considerations such as these regarding the future of cataloging weigh heavily over the enterprise of bibliographic cataloging and classification.

Academic libraries are represented overwhelmingly in the literature from 2009-2010, though many important and useful papers were published about school libraries. Very few articles addressed the unique concerns of cataloging and classification in public and special libraries.

A couple of surprises arose as the author surveyed the literature. RDA was written about in the context of other topics, but few standalone papers analyzing RDA were published. With such a significant change to the fundamental tool used in cataloging looming on the horizon, the author had expected to locate numerous articles addressing RDA. Also, as never before, the business of cataloging was investigated in some detail during the past two years. Most notably, a large study of the MARC records marketplace study was conducted by R2 Consulting for the LC.

The large number of papers on personnel and workflow issues shows that catalogers are still considering workaday activities, but almost all were posited within the context of impending change. The collective output of 2009-2010 ultimately shows that cataloging and classification continue to be essential activities in libraries, even as catalogers partner more and more with those in increasingly-disparate disciplines.

References and Notes

3. Ibid., 1.
4. Ibid.
5. Ibid., 4.
7. Ibid., v.
10. Ibid., 114.
11. Ibid., 113.
13. Ibid., 590.
22. Ibid., 27.
26. Ibid., 10.
27. Ibid., 16.
32. Ibid.
34. Ibid., 1.
39. Ibid., 66.
88. Ibid., 75.
90. Ibid., 21.
92. Ibid., 709.
97. Ibid., 223.
99. Ibid., 200.
107. Ibid., 189.
112. Ibid., 78.
113. Ibid., 200.
117. Ibid., 309.
118. Ibid.
120. Ibid., 223, 226.


128. Ibid., 122.


132. Ibid., 25.