1-1-2008

Knowledge organization

Sherry L. Vellucci
University of New Hampshire, sherry.vellucci@unh.edu

Recommended Citation
Vellucci, S. L. (2008). Knowledge organization. In M.L. Radford and P. Snelson (Eds.) Academic library research: Perspectives and current trends, (pp. 138-188). Chicago: Association of College & Research Libraries, a Division of the American Library Association.

ACRL Publications in Librarianship No. 59

ACADEMIC LIBRARY RESEARCH:

Perspectives and Current Trends

Edited by
Marie L. Radford and
Pamela Snelson

ACADEMIC LIBRARY RESEARCH:
Perspectives and Current Trends

Edited by Marie L. Radford and Pamela Snelson

Association of College and Research Libraries
A division of the American Library Association
Chicago, 2008

Library of Congress Cataloging-in-Publication Data

Academic library research : perspectives and current trends / edited by Marie L. Radford and Pamela Snelson.
p. cm. -- (ACRL publications in librarianship ; no. 59)
Includes bibliographical references and index.
ISBN 978-0-8389-0983-6 (pbk. : alk. paper) 1. Academic libraries. 2. Academic libraries--United States. I. Radford, Marie L. II. Snelson, Pamela.
Z675.U5A3528 2008
027.7--dc22
2008035286

This electronic edition is based on the 2008 printed edition by The Association of College and Research Libraries.

New ISBNs

PDF: 978-0-8389-9541-9
ePub: 978-0-8389-9542-6
Kindle: 978-0-8389-9543-3

4 Knowledge Organization

Sherry L. Vellucci

Since Svenonius analyzed the research base in bibliographic control in 1990, the intervening years have seen major shifts in the focus of information organization in academic libraries. New technologies continue to reshape the nature and content of catalogs, stretch the boundaries of classification research, and provide new alternatives for the organization of information. Research studies have rigorously analyzed the structure of the Anglo-American Cataloguing Rules *using entity-relationship modeling and expanded on the bibliographic and authority relationship research to develop new data models (*Functional Requirements for Bibliographic Records *[FRBR] and* Functional Requirements and Numbering of Authority Records *[FRANAR]). Applied research into the information organization process has led to the development of cataloguing tools and harvesting applications for bibliographic data collection and automatic record creation. A growing international perspective focused research on multilingual subject access, transliteration problems in surrogate records, and user studies to improve Online Public Access Catalog (OPAC) displays for large retrieval sets resulting from federated searches. The need to organize local and remote electronic resources led to metadata research that developed general and domain-specific metadata schemes. Ongoing research in this area focuses on record structures and architectural models to enable interoperability among the various schemes and differing application platforms. Research in the area of subject access and classification is strong, covering areas such as vocabulary mapping, automatic facet construction and deconstruction for Web resources, development of expert systems for automatic classification, dynamically altered classificatory structures linked to domain-specific thesauri, crosscultural conceptual structures in classification, identification of semantic relationships for vocabulary mapped to classification systems, and the expanded use of traditional classification systems as switching languages in the global Web environment. Finally, descriptive research into library and information science (LIS) education and curricula for knowledge organization continues. All of this research is applicable to knowledge organization in academic and research libraries. This chapter examines this body of research in depth, describes the research methodologies employed, and identifies areas of lacunae in need of further research.*

Introduction

In the years since Svenonius analyzed the research base in bibliographic control,[1] there have been rapid and dramatic changes in academic libraries. The resources acquired and made accessible to users appear in many different formats, but increasingly they are digital objects. The term *bibliographic control*, which referred to the traditional organizational functions of descriptive cataloging, subject cataloging, and classification, is principally associated with physical items and is now considered by some to be an inadequate term to describe the range of organizing functions in an increasingly digital networked environment. The terms *information organization* and *knowledge organization* have largely replaced it. While many of the processes remain similar or identical to those which fell under the rubric of bibliographic control, new processes and systems of organization are emerging rapidly.

In this chapter the term *knowledge organization* will be used to include descriptive cataloging, subject cataloging, classification, metadata creation, and the activities of each process that contribute to the making of a catalog or database for the purpose of information retrieval. The term should not be confused with the term *knowledge management*, which has a much broader scope and is primarily found in the context of business environments. The terms *bibliographic data* and *metadata* are often used interchangeably to describe the attributes of a given work; however, the term *metadata* is usually applied in the context of newer organizational systems associated with digital resources. As federated searching becomes more prevalent and metadata creation merges with cataloging functions, the boundaries between the old and new are becoming less clear. These terms, therefore, should not be considered mutually exclusive, as their definitions will no doubt continue to shift with the inevitable evolution of the field of knowledge organization.

In addition to a change in terminology, two noticeable trends have influenced the research and literature of knowledge organization: the continuous development of new technologies and the increasing globalization of information. New technologies continue to reshape the nature, content, and boundaries of library catalogs, the tools we use for information organization, and the work-flow processes. Applied research has enabled the development of computer applications that further automate and assist the information organization processes, in both the library and

commercial outsourced settings. The Internet has allowed the cataloging community to focus on information organization beyond the local and national levels. Thanks to the conceptual, empirical, and applied research in knowledge organization, catalog records can be shared internationally, and global access to digital libraries can be provided. Globalization of information has influenced views on authority control as well and enabled the creation of a Virtual International Authority File. Economies of cooperation and scale necessitate that national libraries, bibliographic utilities, and academic and research libraries collaborate and lead the way in research, development, and implementation of more effective systems of organization.

Another area of internationalization that depends heavily on conceptual research is the ongoing development of international standards. Knowledge organization is inextricably entwined with standards development. Thus, knowledge organization research informs the development of standards, as well as innovations in the processes and systems of organizing information. Cataloging codes and other standards for description and access of all types of resources are now developed in an international context. Currently there are international committees that are helping to redesign the *Anglo-American Cataloguing Rules (AACR2);* the product of their redesign efforts will be called *Resource Description and Access* (RDA). Groups of international cataloging experts are developing a new set of principles and creating and revising new metadata schemas and new ways to structure the data in the catalog. A critical mass of research has been conducted in this area over the past 15 years and has become an important part of the research literature.

Several journals that are specific to knowledge organization publish detailed articles that report on research projects, including *Cataloging and Classification Quarterly, Serials Librarian,* and *Library Resources and Technical Services*. The relatively new publication *The Journal of Internet Cataloging* is devoted exclusively to the organization of digital resources. Other journals that are broader in scope but often include articles on information organization research are *College and Research Libraries; Information Technology and Libraries; Journal of Academic Librarianship; Library Collections, Acquisitions, and Technical Services; OCLC Systems and Services;* and *Technical Services Quarterly.* Shorter articles that discuss research can be found in *International Cataloguing and Bibliographic Control*, *Electronic*

Library, and *Library Hi Tech*. As the boundaries blur between cataloging and metadata, overlap with research in the field of information retrieval (IR) also increases. Because of the vast amount of published research in the area of IR, this chapter will address the IR literature only as it directly relates to library catalogs and related systems of knowledge organization. *Journal of the American Society for Information Science and Technology, Journal of Documentation,* and *Information Processing and Management* are IR journals that often include research articles on knowledge organization, especially in the areas of classification, taxonomies, ontologies, and the semantic web. The extensive amount of research literature in the broad field of knowledge organization necessitates selectivity in the research discussed in this chapter. The author selected the studies discussed here based on the impact of the research on the working academic library environment and the desire to provide an overview of the research areas important to the academic cataloging community.

Academic and research libraries play a vital role in defining the research agenda of knowledge organization, along with the Library of Congress, other national libraries, OCLC, and the Research Libraries Group (RLG). Since academic libraries are often the first organizations to help develop, test, and implement new computer applications and systems, many of these systems have become the de facto standard. The key areas of knowledge organization that will be addressed here include descriptive cataloging, authority control, metadata issues, subject access, and the Online Public Access Catalog (OPAC). Standards are integral to every aspect of knowledge organization, and standards research and development are addressed throughout the text. Although academic libraries are not specifically mentioned in all discussions of the research literature, all of the research examined here has a bearing on academic libraries.

Descriptive Cataloging

Traditionally, the cataloging process is divided into descriptive cataloging and subject cataloging. Descriptive cataloging involves identifying the important characteristics of both the content of a work and the carrier of that content. Another part of the descriptive cataloging process is creating access points for names and titles associated with the work. The creation of access points involves authority control, a process that brings consistency and uniqueness to the access points in a catalog. The second

part of the cataloging process—subject cataloging—involves using controlled vocabularies to assign subject terms and classification notation to describe what the work is about. Both parts of the process should focus on providing an efficient system for users of the catalog to retrieve information. Subject access is discussed later in this chapter.

In the 1990s, several descriptive cataloging issues converged and led to a fundamental rethinking of the conceptual model upon which the *AACR2* were based. These factors included the description of electronic resources with the attendant "content vs. carrier" issue, the growing importance of identifying relationships between works, the ubiquitous access to library catalogs via the Internet, and the desire to share bibliographic and authority data on an international scale. Conceptual research, defined by Svenonius as being "characterized by asking questions, defining terms, imagining possibilities, and analyzing concepts,"[2] was the methodology of choice that enabled experts in the field to rethink and restructure the fundamental concepts of knowledge organization.

Conceptual Models, Theories, and Principles

The most important conceptual research to date has been the development of new conceptual models of the bibliographic universe and the ongoing review and development of cataloging principles better suited to the digital environment. Researchers are examining such questions as "What is a work? How is a work expressed? What kinds of relationships exist among different entities? And given this information, how can we improve catalog functionality for users?"

The Functional Requirements for Bibliographic Records (FRBR)

The IFLA (International Federation of Library Associations and Institutions) Study Group on the *Functional Requirements for Bibliographic Records*[3] (FRBR) began its work with the following objectives:

- to provide a clearly structured framework for relating the data recorded in bibliographic records to the user's needs;
- to create conceptual models for national database systems; and
- to recommend a basic level for national bibliographic records.[4]

In addition, the new conceptual model would be a framework to assist in the development of catalog system designs, in order to take advantage of the computer's ability to link related works in the catalog and

to present a more meaningful OPAC display of the different versions of a work to the user.

An entity-relationship modeling technique was used to develop the new conceptual model. The study group's analysis identified tasks that users might want to perform using a catalog and entities that represent key objects of interest to users of bibliographic databases. The new model represents a shift from a system-centered focus on the functions of the catalog to a user-centered focus on the tasks that catalog users wish to accomplish. The user tasks are to find, to identify, to select, and to obtain. The FRBR conceptual model is composed of three groups of entities. The entities identified in group one are the products of intellectual or artistic endeavor and include the work, the expression, the manifestation, and the item—i.e., the things we catalog. The group two entities are those responsible for the intellectual and artistic content, physical production or custodianship of group one entities, and include persons, corporate bodies, and the recently added entity families. The group three entities serve as the subjects of works and include concepts, objects, events, and places. The links between the entities in all three groups identify the types of relationships that exist between them. The model also identifies the attributes of each entity and maps them to the user tasks served by that attribute.

The FRBR model opened new avenues of research in many aspects of knowledge organization. Some system vendors[5] and bibliographic utilities[6] conducted studies to determine the proportion of works in a particular database that would benefit from application of the FRBR model and to test the viability of applying the model to existing records.[7] Applied research led to the development of new database and record structures,[8] to the creation of record-conversion algorithms to accommodate the model,[9] and to the construction of new interface design tools.[10] Expanding the use of the FRBR model, Naun developed an online journal-finding aid using the FRBR principles to determine the user tasks to be served by the system, the appropriate data structure for the system, and the feasibility of mapping the required data from existing sources.[11] Some researchers are enhancing and extending the model,[12] while others are applying the FRBR model to particular subject domains[13] or reconceptualizing portions of the model.[14] Taniguchi developed a different conceptual model that gives primacy to the expression-level entity rather than the work-level. He

viewed his new construct as a means of dealing with the issue of multiple versions while retaining consistency in the model.[15] In a follow-up study, Taniguchi focused on the whole-part relationship and compared component parts of bibliographic resources using both the FRBR model and his expression-prioritized model.[16] He identified two types of component parts, a "content part" that is contained within the physical structure of the host resource and a "document part" that is physically separate. He concluded that in the FRBR model, the "whole" and the component part are modeled in the same way, but a different model is used for the content part. In Taniguchi's model, all three are modeled in the same way (the whole resource, the component part, and the content part), thus, in his opinion, making the model more consistent than FRBR. Considering the problems that the FRBR model now has with aggregated works, it is possible that by giving primacy to the expression level, Taniguchi's model would solve this problem.

The FRBR model has had an impact on the cataloging rules used in the United States and abroad. A study by Delsey and others rigorously analyzed the structure of the *AACR2* using the entity-relationship modeling technique. "The principal objective of this study [was] to develop a formalized schema to reflect the internal logic of the *Anglo-American Cataloguing Rules,* [which in turn would] serve as a tool to assist in the re-examination of the principles underlying the code and in setting directions for its future development."[17] This analysis was the preliminary step in what is to be a new content standard for bibliographic records. Envisioned for use beyond the cataloging community, the new cataloging standard, the RDA, will be independent of record structures (e.g., MARC 21) and will integrate the FRBR conceptual model and its terminology.[18] Major criticisms from important constituent groups regarding, among other issues, the lack of FRBR integration have compelled the developers to abandon the first two drafts of RDA. The RDA standard is scheduled for completion in 2009; however, many believe the reconceptualization of the standard will meet the needs of neither 21st century catalog users nor the broader metadata communities as was hoped.

While it is too early for a substantive body of RDA research to appear, Dunsire explored the basic concept of separating content description from carrier description in order to develop the RDA/ONIX Framework for resource categorization.[19] The framework, which is designed to improve

metadata interoperability between libraries and publishers, identifies and defines two distinct sets of attributes: (1) the intellectual or artistic content of an information resource and (2) the type of carrier for such content. The framework provides for constructing higher-level categories of resource content and carrier from the attribute and value sets and includes recommendations on applying such categories to resource descriptions.

Several research projects have focused on mapping the FRBR entities and attributes to other standards. Delsey produced a research report for the IFLA International Standard Bibliographic Description Review Group (ISBD) that analyzed and mapped "each of the elements specified in the ISBDs to its corresponding entity attribute or relationship as defined in the FRBR model."[20] LeBoeuf examined the impact that the FRBR model would have on future revisions of the ISBD standards.[21] Delsey also conducted a conceptual analysis of the MARC21 (Machine-Readable Cataloging) communications format and mapped the FRBR data elements to the MARC 21 data elements.[22] The mapping was updated and revised in 2004 by the Network Development and MARC Standards Office at the Library of Congress (LC).[23] All of these critical analyses and mappings help to clarify the FRBR entities and data elements by placing them within the context of standards that are already familiar to catalogers.

Works and Relationships

Empirical research uses quantitative and qualitative methods to measure and analyze existing phenomena and is dependent on appropriate constructs and interpretation of the data to inform future research and decision making. The FRBR conceptual model draws on a body of empirical research that examined the nature of a work and the concepts of bibliographic families and bibliographic relationships. The identification and referencing of bibliographic and authority relationships is a formal way to create the syndetic structure of the catalog. Research in this area is ultimately focused on finding better ways to exploit the capabilities of computers to provide a more meaningful grouping for a clearer presentation of related entities described in a catalog. The seminal research conducted by Tillett used both analytic and empirical methodologies to examine bibliographic relationships in depth.[24] Her study began with the creation of a taxonomy of bibliographic relationships that she discovered by examining 24 different cataloging codes from Panizzi's 1841 *Rules* to the

1978 edition of *AACR2*. The results of her analysis identified seven types of bibliographic relationships, which consist of equivalent, derivative, whole-part, accompanying, sequential, descriptive, and shared characteristics. Tillett also examined the cataloging codes to identify the various types of linking devices used to establish each type of relationship on the bibliographic record. The second part of Tillett's work included an empirical study designed to examine the extent of bibliographic relationships as reflected in their frequencies of occurrence in MARC records entered in the LC machine-readable database between 1968 and 1986. Although there were problems with the sampling frame due to the types of materials cataloged by LC in the MARC format at that time, one important finding of this portion of the study indicated that bibliographic relationships were widespread throughout the bibliographic universe, i.e., Tillett found that almost 75% of the records in the database contained some type of relationship information.[25]

Tillett's landmark study became the starting point for two further investigations that focused on the bibliographic universe represented in library catalogs, both of which were narrower in scope. Smiraglia conducted an empirical study of works that focused on the derivative relationship, one of the most frequently found relationships in the catalogs of large academic libraries.[26] His analysis identified seven types of derivative relationships found in 49.9% of works in his sample. The derivative types identified by Smiraglia include simultaneous derivations, successive derivations, translations, amplifications, extractions, adaptations, and performances. Smiraglia's research findings indicate the importance of identifying and linking surrogate records for members of bibliographic families in academic library catalogs. Further research by Vellucci examined the bibliographic universe of musical entities to identify bibliographic relationships found in catalogs representing large collections of music.[27] Vellucci's sampling frame used the musical scores catalog of Sibley Library at the Eastman School of Music. She then searched the OCLC and RLIN (Research Library Information Network) databases to identify bibliographic entities related to each member of the sample, then analyzed and categorized the types of relationships found to exist for musical works. Her findings indicated that 97% of the music scores in the sample exhibited at least one relationship, a considerably higher figure than that discovered by Tillett in the general bibliographic universe. Vellucci's research findings

suggest that the overall proportion of relationships found among entities in a catalog may differ by discipline and format of the entity. This research has direct implications for OPAC transition to a new FRBR-based system design, for the results indicate that a large portion of works represented in a music catalog would benefit from use of the FRBR model.

In addition to identifying types and degrees of relatedness, all three studies examined the methods by which relationships were identified and represented in library catalogs at the time of the respective studies. Each study concluded that new OPAC and surrogate record designs were necessary to enable better exploitation of the computer's capability to link and display related entities in library catalogs. A thorough analysis of the history, research, issues, and contexts of bibliographic relationships was presented by Vellucci at the Toronto Conference on the Principles and Future of AACR.[28] The paper concludes with her identification of four fundamental principles for the treatment of bibliographic relationships that should guide the development of cataloging codes and system design. These are the principles of relationship identification, enabled linkage, multilevel description, and consistency.

Yee focused her research specifically on the concept of the work.[29] Drawing on her doctoral research, she showed the lack of a formal definition of a work throughout the history of the *AACR2* codes and concluded with an extensive definition of a "work," which evolved from her research. Included in her definition are the separate concepts of expression and publication, which are later used in the IFLA FRBR model as the expression and manifestation entities. Carlyle's dissertation research examined how works were collocated in OPAC displays and suggested ways to improve the grouping of bibliographic families.[30] Her later research continued to focus on improving catalog displays through grouping related resources by investigating how users categorize works.[31] In another study, Carlyle conducted a survey of 18 online catalogs to compare the displays resulting from five author and five work queries.[32] A more recent article by Carlyle and Summerlin discussed their research on record clustering of works of fiction to improve catalog displays.[33]

Continuing his earlier research, Smiraglia examined the work entity in greater depth.[34] His theoretical analysis drew on linguistic and semiotic theories to develop a new "theory of the work." Smiraglia's concept of the work has greatly influenced the ongoing refinement of the FRBR model

and conceptual models developed in other communities.[35] In order to estimate the number of works in OCLC's WorldCat database, Bennett, Lavoie, and O'Neill constructed a sample of works by applying the FRBR model to randomly selected WorldCat records.[36] This sample was used to describe the key characteristics of works. Results suggest that the majority of benefits associated with applying FRBR to WorldCat could be obtained by concentrating on a relatively small number of complex works. Finally, a special issue of *Cataloging and Classification Quarterly,* edited by Smiraglia, was devoted to modeling a wide variety of works, including fiction, television series, videos, digital resources, and cartographic resources, as well as scientific, multimedia, collected, and theological works.[37]

The generalized nature of the FRBR model is both a strength and a weakness. Its strength is as a logical framework that provides common ground for further discussion and research on bibliographic data and the entities that these data describe. Another strength of the model "lies in its separation of the logic and principles of description from display issues."[38] Problems with the FRBR conceptual model arise with its implementation, for the generalized entities, attributes, and relationships do not provide enough detail on which to develop a database. Another weakness of the model lies in the method used for identifying user tasks and mapping entity attributes to specific user tasks, i.e. critical analysis at a high level. This research calls into question the validity of the user tasks identified and used in the model. User studies are needed to confirm the validity of these user tasks and to test the accuracy of the data mapping by comparing both to the real world within the context of the catalog and the bibliographic record data elements that satisfy specific user information needs. As the IFLA Study Group on FRBR identified a wide range of catalog users both within and outside the library environment, the user studies must be conducted with similarly wide user groups. In addition, the report gives no evidence that the body of research on information-seeking behavior was consulted, much of which would be useful in confirming the validity of the users' needs.

Problematic issues with the FRBR model continue to be examined.[39] One such problem is that of aggregate entities, which becomes particularly fuzzy when dealing with serial works. Antelman's analytical research examined serial work identifiers used by the rights-holder community (e.g., International Standard Serial Number [ISSN], Digital Object Identifier [DOI], etc.) and dismissed their usefulness as work identifiers in the

library community. She notes that "there is a practical need in bibliographic control for a level of abstraction that brings together related items that do not exhibit textual identity,"[40] citing other terms such as *superwork, superwork record set, super records,* or *package content* that have been used by other researchers. Antelman then developed a conceptual model of a serial work based on the FRBR model, giving primacy to an abstract work level that collocates the bibliographic family, followed by her concept of different expressions of the work. Her manifestation level contains the published versions of the expressions. Included at the manifestation level are different aggregator versions of selected articles. The model is not logically consistent, as it conflates the manifestation and item levels for nondigital versions of the serial (paper, microfilm, etc.) and separates the digital versions into a manifestation level consisting of aggregator versions of separate articles and an item level comprised of the different file formats (HTML, PDF, ASCII, etc.). Nor does the model consider the possibility of a serial as a "work of works," with an issue or article as a work in its own right. Antelman's research is useful, however, because it brings a new perspective to modeling serial works, but the question "What is the work?" is still not resolved for serials. Flexibility seems to be the answer at this point, but further research that includes different serial models is needed in order to determine how to handle aggregates and identify the problems that would arise when trying to accommodate this flexibility in a catalog.

Hirons and Graham developed a conceptual model for seriality that strongly influenced the revision of the *AACR2*.[41] Their model created a three-dimensional approach to cataloging serials that is based on the attributes of the content, the carrier, and the publication status of the work. Static materials are those that are complete when issued. Ongoing or continuing resources are those that are not complete when first issued; these can be either indeterminate or determinate. Determinate publications are continuing resources that are intended to be complete in a finite number of parts or over a finite period of time. Indeterminate publications are continuing resources that are intended to continue indefinitely.[42] Resources in both of these categories can be multipart or single-part updating. Another layer is added to the multipart indeterminate resources, which can be either numbered or unnumbered. This model is especially useful when dealing with looseleaf materials and electronic serials.

Finally, an important new publication edited by Taylor provides a general overview of FRBR and offers chapters by experts in the field that examine the FRBR model in relation to a wide variety of topics, including the Functional Requirements of Authority Data (FRAD), the history of cataloging, research, bibliographic families, RDA, archival materials, moving image materials, music, and serials.[43]

Another important conceptual research project being conducted by IFLA is the creation of a new set of principles for the development of international cataloging codes.[44] Building on the conceptual model for the FRBR and the draft of the FRANAR (see below), IFLA has conducted several International Meetings of Experts (IME) for input into the process. "These new principles replace and broaden the Paris Principles [of 1961] from just textual works to all types of materials and from just the choice and form of entry to all aspects of the bibliographic and authority records used in library catalogs."[45] IFLA intends for these principles to be applicable to online library catalogs and other appropriate databases. Also included in the new statement are objectives for the construction of cataloging codes.

Electronic Resources

Many studies have focused on the description and access of special types and formats of materials. The largest body of research in this area over the last decade concentrated on the newly emerging electronic resources.

In the 1990s, the need to describe digital objects challenged the adequacy of *AACR2*'s cardinal principle—to describe the item in hand. This tradition of description based on physical format created obstacles for dealing with issues such as multiple versions and electronic resources. The increasing complexity of the bibliographic universe called into question the role of the catalog and the ability of the current cataloging code to describe electronic resources in a networked environment. In 1998, OCLC began development of the Cooperative Online Resource Catalog (CORC), a major research project to test the process of cataloging electronic resources using traditional library standards and newly developed software applications.[46] When the prototype system went online in 1999, academic and research libraries were major participants in beta testing the system, eager to try new methods of providing access to electronic resources. The CORC prototype incorporated several software applications developed by

OCLC, including Mantis, Kilroy, Scorpion, and Wordsmith.[47] Traditional cataloging standards were used, including *AACR2*, the MARC format, *Library of Congress Subject Headings (LCSH)*, *Library of Congress Classification (LCC)* and the *Dewey Decimal Classification (DDC)*. The research project also included experiments with reciprocal conversion of MARC records and Dublin Core metadata records, automatic assignment of subject headings, automatic assignment of *DDC* numbers, and automatic creation of metadata extracted from the Web site.[48] During the beta-testing period, research by Hsieh-Yee and Smith indicated that while overall the participants considered the experience positive, there was need to improve the speed of the system to make it a viable working tool.[49]

At the end of the CORC experimental stage, Connell and Prabha conducted a study using a proportional sample of member-created records in the CORC database to examine characteristics of the resources represented.[50] The results indicated that academic libraries were the largest contributors to the database, adding 67% of all records. This finding supports the notion that academic libraries have a high level of concern about providing access to electronic resources for their users. One unexpected result from this study showed that only 21% of the CORC records were for resources held locally by the contributing library, with 78% being external to the institution. Although the researchers expected libraries to be more concerned with making their own unique resources available on the Web, upon consideration, they decided that the likely explanation for this outside focus was that the time and expense to create digital versions of their unique materials would be much greater than adding records for existing Web resources. This hypothesis needs to be verified by further research. The researchers also categorized the Web resources by discipline, publication pattern, and the unit being cataloged. It was noted that current definitions for Web units were inadequate for categorizing the level of granularity for a Web site and further research was needed to develop definitions that would be meaningful to users.

Additional research conducted during the CORC Project studied the potential for automatic classification and description of Internet resources and examined options for catalog interface displays.[51] OCLC developed the Persistent URL (PURL) for the CORC project to alleviate a library's burden of periodically checking each record to ensure that the URL links were active.[52] Another tangible result of this research project was the revi-

sion of the MARC 856 field to accommodate more electronic resources information and the revision of the MARC 856 subfield ≠u to end repeatability of the URL and require each location of an electronic resource to be entered in a new 856 field. Records created in this research project have been added to the WorldCat database and contributed a substantial group of records for Internet resources.

During and following the CORC project, academic librarians began to experiment with different methods of describing and accessing the different types of electronic resources identified above by Hirons and Graham. The following studies were selected from a large body of literature in this area.[53] At Cornell, Calhoun explored a new team-based model that required crossfunctional collaboration. In this model, data for description and access could be gathered from selectors, public service librarians, information technology staff, authors, vendors, publishers, and catalogers.[54] Because cataloging electronic resources was a new issue for most libraries, several studies applied survey research to gather descriptive information about local practices. Chen and others surveyed academic libraries to discover their cataloging practices for electronic resources.[55] Their findings show that the task of organizing electronic resources presented librarians with a host of new and complex challenges. "This volatile set of unstable resources … change names, contents, providers and URLs with alarming frequency … [requiring] repeated revisions to their surrogate records."[56] This was complicated by the lack of comprehensive standards for cataloging electronic resources. Chen and others made several other discoveries, including the fact that all libraries were presenting holdings information on Web lists instead of, or in addition to, the OPAC; there was no consistency on whether to catalog different formats on individual records or one integrated record; there was no consistency on which part of the resource a URL linked to; and volatile URLs were difficult to cope with. Trends in cataloging electronic resources were also identified. Martin and Hoffman conducted a similar study that focused on cataloging journal titles in aggregator databases.[57] Li and Leung discussed the development of a software program to automatically integrate full-text electronic journal titles in unstable aggregator databases into a library' s OPAC.[58] Banush, Kurth, and Pajerek developed an automated system for controlling serial titles in the catalog.[59] A large portion of the literature on organizing electronic serials is case studies and descriptive analyses

of problems and solutions representing practice and opinion. Copeland provided a review that discusses much of this literature in detail.[60]

Authority Control

The FRANAR IFLA Working Group was charged with developing a conceptual model for authority records and studying the feasibility of an International Standard Authority Data Number (ISADN). Continuing on the work initiated by the FRBR Study Group, the FRANAR Study Group's research and analysis identified five functions of an authority file (document decisions, serve as reference tool, control forms of access points, support access to bibliographic records, and link bibliographic and authority files) and four tasks important to users of authority records (find, identify, contextualize, and justify). The initial draft reports containing the conceptual model were released for worldwide comment under the title *Functional Requirements for Authority Records* (*FRAR*); the latest draft has been retitled *Functional Requirements for Authority Data* (*FRAD*)[61]. The FRAD conceptual model expands on the FRBR Group Two entities (person and corporate body) to include the new entity *family*. The model then identifies attributes and relationships of the three FRAD entities and the FRBR Group One entities (work, expression, manifestation, and item) within an authority record context and maps each attribute to a specific user task. A new IFLA Working Group will analyze the "entities that are the centre of focus for subject authorities, thesauri, and classification schemes, and of the relationships between those entities," and develop a conceptual model of "Group 3 entities within the FRBR framework as they relate to the aboutness of works."[62] On a final note, all of these IFLA reports thus far claim to be user-centered, yet no research was conducted involving users, and the user tasks were based on educated assumptions, thus resulting in a lack of validity and generalizability for those aspects of the reports. This lack of user research does not, however, detract from the importance of the conceptual models in clarifying and mapping the complex bibliographic universe.

In addition to these major IFLA research projects, other authority control research in descriptive cataloging continued to examine more specific segments of the topic. Much of this research falls into the categories of retrievability, including international issues, and process and cost issues. The measure of success and the ultimate goal for authority control

research is retrievability. Studies in this category examined data quality, automation, foreign and nonroman languages, and internationalization. Using a transaction log analysis methodology, Taylor's research examined variations appearing in name access points found in OCLC bibliographic records.[63] In a similar study, Weintraub examined the effects of personal name variations on authority control and data retrieval in computerized catalogs.[64] Bowman examined 38 OPACs available on the Internet in order to identify retrieval problems resulting from lack of crossreferences and variant author names.[65] Jin compared forms of corporate names found in the Library of Congress Name Authority File with the forms found on corporate Web sites.[66] Calhoun and Oskins used several different variables to investigate the types of changes made to LC authority records over time.[67]

Several studies addressed the authority control problems dealing with foreign and nonroman names. Ruiz-Pérez and others investigated how Spanish names were handled in three different databases and identified mistakes that would have a negative impact on locating and retrieving works by Spanish authors.[68] Hu used a transactional analysis method to examine problems in cataloging Chinese names, including transliterating Chinese characters and establishing entries for personal names.[69] Bolick also studied Chinese names, examining the reasons why vendor software produced negative results for nonunique Chinese names.[70] Wang examined the problems of Chinese names in different languages, i.e., English transliteration, transliterated name from spoken Chinese dialect, westernized Chinese, and pure Chinese.[71] Influenced by a Machine-Readable Bibliographic Information (MARBI) discussion paper on multilingual authority records,[72] Plettner developed three models for entry of Arabic names.[73]

Authority control has always been the most time consuming and expensive part of the cataloging process, so improving the quality of the data for retrievability, while reducing the time and labor costs involved with the process, is an ongoing area for research. Research in the process and costs category included quality control and evaluation research, workflow studies, and cost-benefit analysis. Kulczak used the OCLC database to evaluate the quality of front-end authority work for monographs in order to determine whether authority work was necessary at the copy cataloging stage.[74] Pappas conducted a similar study using records from the RLIN database.[75] Beall investigated the impact of typographical errors

in authority records.[76] Ellero investigated the use of the Web as a source of information for authority records.[77] Jin adopted a narrower research focus and examined the process of constructing corporate name headings using data from the corporate body's home page.[78] Wolverton surveyed the authority control practices and staffing and training issues related to authority control in large academic research libraries.[79] Bangalore and Prabha investigated the copy cataloging process by measuring the time and effort expended.[80] Santizo and Rezabek surveyed academic libraries to find information about the type of decisions and level of responsibility of authority control required of copy catalogers.[81] Tsui and Hinders conducted a cost-benefit analysis for outsourcing the authority control process.[82]

Technology and automation were important catalysts for authority control research, much of which used applied and experimental research to develop new software tools and systems and to integrate authority records with bibliographic records in databases. The primary goal of this type of research is to reduce the amount of human intervention needed to create authority records. One method of evaluating these automated systems is to measure the accuracy of the system in terms of its ability to disambiguate like names and titles. As development of the Virtual International Authority File (VIAF) progresses, this research is particularly concerned with merging large data files. Merged records in the VIAF contain the form of name or title used in different countries. Local libraries will have options when deciding which form or forms to display. Another evaluation method is to test how well an automated system performs when its accuracy is compared to a human-created authority record. DiLauro and others presented a detailed report of the process used to create an automated authority record for the Lester S. Levy Collection of Sheet Music.[83] Patton and others developed and tested an automated name authority control (ANAC) tool as part of the Lester S. Levy Sheet Music digitization project.[84] The goal of their research was to reduce the costs associated with applying authority control to the metadata for the collection, and "the development of tools that combine automated processes and human intervention, with the overall goal of involving humans at higher levels of analysis and decision making."[85] The researchers extracted names from the statements of responsibility in their metadata records using a rule-based name extraction algorithm developed for the project.

The resulting name records were automatically compared to the records in the LC Name Authority File, using probability theory to determine a match. "Overall, ANAC was successful 58% of the time. When a name had an LC record, ANAC was successful 77% of the time, but when an LC record did not exist for a name ANAC was successful only 12% of the time."[86] These results confirm that the ANAC system would not function well without human intervention in some cases. The researchers emphasized that their intention was not to eliminate human intervention completely from the process, but to be able to predict the amount and type of intervention that would be needed.

In order to test the dependability of a new authority control module in the library's automated system, Greever compared the module's effectiveness to the currently used manual procedures for authority control.[87] Running the two systems simultaneously, Greever compared the number and type of established headings for which authority records would be added to the local system using the precataloging procedures to the number and type of established headings for which local authority records would be added by the automated system. Her results showed that the new automated module was in fact equal to or more effective in establishing headings and the new system reduced redundancy and generated fewer errors. French and Powell investigated approximate string-matching techniques and introduced the concept of approximate word matching to improve detection and categorization of variant forms of names.[88] Miller discussed the development of XOBIS, an XML schema that reorganizes bibliographic and authority data elements into a single integrated structure.[89] The LEAF (Linking and Exploring Authority Files) research project is developing a model architecture for collecting, harvesting, linking, and providing access to existing local or national name authority data.[90] The architecture will provide a mechanism to search the authority records of individual data providers and merge the information into one LEAF authority record containing international name data that will be stored in a "Central Name Authority File."

In addition to these experimental and developmental studies, the literature contains examples of local implementation reports on the topic of outsourcing authority control work. Lam described the outsourcing experience of the University of Saskatchewan.[91] This experiential report included statistical information on each phase of the process and cost

information for an initial database clean-up, an annual authority database updating, and a weekly service for current authority work. The university was satisfied with the vendor's overall linking rate of 85.9% for headings. An analytical study conducted by Vellucci describes the outsourcing process in detail and speculates on the future problems of outsourcing authority control on an international scale.[92] All of this research provides useful information for practical decision making. It is important, however, to weigh the results of any given study against a comparable situation in terms of the type of library, how a system functions, the nature of the records being processed, and the participants in the research before generalizations can be made.

Finally, a major international conference on authority control was held in Florence, Italy, in 2003, and proceedings were published.[93] Forty-eight papers were presented, organized into the following categories: state of the art and new theoretical perspectives; standards, exchange formats, metadata; authority control for names and works; authority control for subjects; and authority control experiences and projects. The proceedings provide a wealth of information on every aspect of the topic, including research using a variety of methodologies.

Metadata

Metadata is structured data that describes a resource, identifies its relationships to other resources, and facilitates the discovery, management, and use of a resource. Although traditional cataloging information can accurately be described as metadata, most people use the term in the context of digital resources. A major paradigm shift occurred in knowledge organization with the development of metadata schemas and markup languages that could serve as alternatives to the *AACR2* and the *MARC Bibliographic Format* for catalog records. Both cataloging and metadata require the same conceptual knowledge; it is the application and management of the metadata that differs. The processes of creating metadata and cataloging are so closely linked that in many academic libraries the boundaries between the two are obscured and the two units are often merged into one department. This merger opens new opportunities for academic library catalogers to expand their expertise. The research conducted in the area of knowledge organization is also becoming more integrated, and the work that is done in one community often has an

impact on the other. Due to the large number of studies conducted in the area of metadata, the research discussed in this section is selective and will provide only a general overview of the directions in which metadata research is progressing.

Several general metadata schemas exist such, as the simple and popular Dublin Core Metadata Element Set (DCMES) and the more complex Encoded Archival Description (EAD) and Text Encoding Initiative (TEI). In addition to these general schemas, many subject-specific metadata schemas have been developed since the mid-1990s; however, discussion of each metadata set is beyond the scope of this chapter.[94] A great deal of conceptual research was invested in developing these schemas; those efforts are reflected in the descriptive nature of the research that was initially published. As the field of metadata matured, researchers looked beyond metadata schema description and began to examine a wide range of metadata related issues. These ongoing research initiatives fall into several categories that are not mutually exclusive, including options for metadata creation, development of tools to aid metadata creation, expanding metadata interoperability, developing vocabularies to support the semantic web, and developing ways to evaluate the quality and usability of metadata.

Metadata Creation

Because of the substantial number of digital resources requiring metadata for description and retrieval and the high cost and subjectivity of human-created metadata, several researchers explored alternative methods for the metadata creation process. One option is to have author-generated metadata. Since the Dublin Core was originally designed for resource authors to create their own metadata, in a baseline study using a mixed-mode methodology, Greenberg and others examined the ability of authors to generate acceptable Dublin Core metadata for their own resources.[95] The study involved a training session for the authors prior to metadata creation, author creation of metadata using a template designed for the project, analysis of the newly created metadata by a research team of experts, and a survey questionnaire to ascertain the authors' perspective on the project. The results of the study indicated that 64% of the metadata records were either good or excellent and 36% needed major revision. All records were considered acceptable for resource description and retrieval

by the research analysis team. From author questionnaire feedback, the researchers realized that a redesign of the input template was necessary. After the redesign, the study was conducted again with 29 participants, and the authors were asked to answer additional questions about their metadata experience.[96] These findings showed serious problems with author motivation for creating metadata themselves. Several authors believed that it was not their job or interest to create metadata, and others questioned the benefits of metadata. The study concluded that it is imperative to create metadata input tools where the interface template is clear, is easy to use, and provides a comprehensive explanation of the metadata process. Taking into consideration the small sample size of both studies and the fact that the studies were conducted in an organizational environment, the results are not generalizable to the academic setting. Academic libraries, however, might wish to replicate this study with faculty and noncataloging staff to determine if the results are comparable, if the quality success rate is acceptable in academe, and if there are institutional factors that might influence metadata creation.

Wilson conducted a recent study of contributor-supplied metadata[97] using the RILM Abstracts of Music Literature (RILM), an international database of scholarly works about music.[98] The RILM database relies on authors and subject-expert volunteers to provide the basic metadata records and abstracts that describe the literature. Metadata records submitted through a Web form were compared to the final published version of the record that had been reviewed and enhanced by a metadata professional. A sample of English-language records were examined for quality based on the completeness of each record; the types of errors (typographical/grammatical or semantic), if any, in each record; and the appearance and type of "value-added" or additional metadata supplied by the contributor in each record. Structural and semantic errors were noted throughout the records evaluated. Results of the study showed that overall, the semantic quality was very high, reflecting the subject expertise of the contributors. The appearance and type of structural errors suggested that improvements to the interface of the Web form, such as drop-down lists, could reduce structural errors to create high-quality contributor-supplied metadata records.

Another option for metadata creation is to develop applications that can automatically extract data from the document and generate metadata.

Earlier reference was made to the Johns Hopkins University's Levy Digital Sheet Music Project experience with automatically generated authority records.[99] A sample of other research work will be mentioned here. Yilmazel, Finneran, and Liddy developed a natural language processing system to automatically assign metadata.[100] A collection of educational documents was used, and three distinct extraction modules were designed to compile the data, along with constant data extracted from the collection level file. The results of the data extraction from all modules were then collocated, prioritized, and output as a single metadata record. The generated and manual metadata records were then evaluated by teachers for how well the metadata represented the lesson plans. The results indicated that in most data elements there was no significant difference between the manual and generated metadata. The only elements where the manual metadata were significantly better were title and keyword.

Paynter described a large ongoing project to develop automatic metadata assignment and evaluation tools for the INFOMINE Project.[101] "The assignment tools that resulted from this research range in complexity from simple rules for assigning Title and Creator metadata by harvesting the text of HTML tags, to Keyphrase and Description extraction algorithms based on syntactic processing of Metatags and to complex Library of Congress Classification (LCC) and Library of Congress Subject Heading (LCSH) classifiers based on algorithms."[102] Manually assigned metadata was used to train the machine learning tool for assigning both LCC and LCSH metadata, and an automatic evaluation tool was used to determine the metadata quality. Paynter included extensive discussions of the extracting and assigning processes for each data element and a useful survey of the research previously conducted for each method used.

All of these studies indicate that automatically generated metadata, combined with manually created metadata for specific elements, can optimize the quality of the metadata while reducing the time and cost of having experts create the entire metadata record. Academic libraries could use this information to create a new metadata creation model that would be especially useful for supporting digital library collections and metadata repositories.

Common Research Goals

As mentioned earlier in this chapter, the overlap between cataloging and

metadata has blurred to the extent that the boundaries are often artificial. The research discussed in this section is of mutual concern to the cataloging and metadata communities. Globalization, economies of scale and cooperation, the plethora of knowledge organization systems (KOS), and the advancement of technology are all factors leading to a confluence of research among the cataloging, metadata, and information-retrieval domains. Key problems common to all include interoperability and automatically generated subject terms and classification.

Interoperability

The American Library Association has defined interoperability as the ability of two or more systems or components to exchange information and use the exchanged information without special effort on the part of either system.[103] Interoperability research extends to developing crosswalks between different KOS, searching and retrieving data from multiple repository databases, simplifying subject access, mapping subject terms among different vocabularies and languages, using classification as a switching language, and developing taxonomies and ontologies for specific subject domains. This last area of taxonomies and ontologies will not be covered in this chapter.

Technical Interoperability

Interoperability is an important issue for academic libraries that participate in distributed networked environments where cross-domain, cross-repository, and cross-language searching are increasingly important. Interoperability research can be viewed from two perspectives: technical data transfer and/or conversion and content compatibility. The research on the technical transfer and conversion of metadata focuses on system architecture, record structure, syntax, and types of data elements. Content interoperability research is concerned with standards for description, such as *AACR2R* or the *International Standards for Archival Description* (ISAD), authority control for names, identifiers, controlled vocabularies, natural language processing and classification. Both aspects of interoperability are critical for the efficient exchange of high-quality metadata with or without human intervention.

Crosswalks are used to facilitate machine processing and data transfer. "A crosswalk is a mapping of the elements, semantics, and syntax from

one metadata schema to those of another."[104] The elements (or fields) in one metadata set are correlated with the elements of another metadata set that have the same or similar meanings. Many metadata schemas are mapped to the MARC format, using crosswalks to exchange and convert data to and from MARC. Conceptual research is used to develop a single file table of equivalency, while applied research is the basis for crosswalk implementation. Problems arise when metadata records with incompatible data elements in the two schemas are mapped to each other. It is often the case that there is not a one-to-one match between all data elements in the two schemas and data from the richer schema is lost during the transformation process. In addition, as these single files contain all information in one equivalency table, each crosswalk is limited in its use and usually applies only to a single digital project.

A research project conducted at OCLC by Godby, Young, and Childress addressed this problem by developing a repository of metadata crosswalks.[105] The project goes beyond the standard single file mapping table and uses the Metadata Encoding and Transmission Standard (METS), a structural framework, to define a complex data model that brings together three separate files: one for the table of equivalence, one for the source metadata standard, and one for the target metadata standard. These data are then processed to create XML-encoded METS records for each crosswalk and are available for processing by "search engines, OAI (Open Archives Initiative) harvesters, and custom-designed web services."[106] By using separate files to maintain the element sets of each metadata schema and each equivalency table, one can create a customized crosswalk by selecting the metadata and table of choice. This prototype is an important step forward to improve and standardize crosswalk formats; however, the authors note that this area of research must continue in order to develop "robust systems that handle validation, enhancement, and multiple character encodings and allow human guidance of the translation process."[107]

The OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting) is an applied research project that seeks to develop an efficient system to search and retrieve metadata simultaneously from multiple database repositories. Several years of conceptual and applied research went into "drafting and designing a useful and useable technical specification" that is now in use by many academic libraries.[108] Differing from other protocols, the OAI model was designed to collect metadata by harvesting

(i.e., "pull") rather than contribution (i.e., "push") to one central repository. After the initial implementation, the model was refined to extend its capacity to identify and harvest a variety of metadata schemas using the Dublin Core as the mandatory standard. An additional refinement of the protocol model allowed for discovery and retrieval of data in repositories of nondigital objects.

Van de Sompel and others identified several limitations with the OAI-PMH protocol.[109] The Dublin Core element "dc.identifier" was not expressive enough to distinguish among the variety of identifiers permissible in the Dublin Core and therefore could not accurately harvest the actual digital resources. The inability to harvest the resources described by the metadata was also a problem for their preservation, which requires archiving a large, synchronized repository. Van de Sompel and others designed a new model to extend the capabilities of the OAI-PMH to include harvesting resources as well as metadata. The model represents increasingly expressive metadata formats contained in a XML document wrapper. The first level below the actual resource is the OAI-PMH identifier, which acts as the entry point to all of the metadata formats. These formats include the minimalist Dublin Core Metadata format, the richer and more expressive MARCXML metadata format, and the METS XML complex object metadata format, which is the most complex and accurate description of the resource. This combination of formats allows for greater depth of expressiveness and provides secondary information that was not accessible in the original OAI-PMH model, including rights management and technical, structural, and provenance metadata. The new model also provides access to the resource, thus solving the problems of unambiguous resource discovery and harvesting of resources for preservation.

The importance of the new model becomes evident in a federated search environment. The model allows academic libraries to reuse their MARC metadata and thus provides integrated access to metadata for nondigital objects and allows harvesting of digital resources for access or preservation purposes.

Subject Access

As the need for internationalization and the volume of Web resources grew, there was heightened interest in improving the interoperability of KOS and making these systems easier to use. Solutions to these goals are

not limited to any one domain, and the literature of cataloging, classification, metadata, and information retrieval all provide examples of research in this area. The overlap of research interests has created a vast body of literature, and of necessity the studies discussed here are selective examples to provide some understanding of the breadth of research. Researchers are investigating the development of less complex subject systems, automatic generation of subject headings and classification, multilingual subject access, and tools that will aid the user in subject searching.

In response to the trend to simplify cataloging and reduce costs by eliminating controlled vocabulary terms from catalog records, Gross and Taylor conducted research to identify the proportion of OPAC records retrieved by a keyword search that had the "keyword only in a subject heading field and thus would not be retrieved if there were no subject headings."[110] The authors used a transaction log analysis methodology to sample keyword searches in an academic library catalog and then performed keyword searches in an OPAC using those terms. Gross and Taylor found that an average of 35.9% of records retrieved by successful keyword searches would not be retrieved if subject headings were not included in the record. Their research also discovered many individual cases in which 80, 90, and even 100 percent of the retrieved records would not be retrieved in the absence of subject headings.[111]

Zeng and Chan reviewed the research methodologies used in projects to establish content interoperability among KOS that focused on subject terms and classification schemes.[112] The authors identified several problematic issues for subject interoperability, including mapping systems with different structures and characteristics (e.g., controlled word-based vocabularies to classification systems), mapping similarly structured schemes (e.g., word-based vocabularies to other word-based vocabularies, or classification schemes to other classification schemes), mapping vocabularies in different languages, and integrating the views of different cultures. Zeng and Chan identified several mapping methodologies, among which are derivative modeling (creating a simpler KOS based on a more comprehensive system), direct mapping (equivalence matching term to term or term to classification), translation and adaptation, and using a switching language to mediate between languages. The authors report on research projects dealing with each of these issues and methodologies.

The exponential growth of electronic resources led researchers at OCLC to seek ways to provide subject access to large volumes of resources, yet make it easier and less costly to use than the complex syntax of the LCSH.[113] Implementations of the minimalist Dublin Core metadata schema created the impetus to develop a subject vocabulary that was simple in structure, provided optimal access points, and was flexible and interoperable across a variety of KOS, including OPACs. Deriving their new vocabulary from the LCSH and simplifying the syntax, the researchers created the Faceted Application of Subject Terminology (FAST) as a postcoordinate approach. Literary warrant was used to determine which headings would be established, based on the frequency with which a heading appeared in the OCLC WorldCat database. FAST headings were established by deconstructing the LCSH headings into facets, which enabled the vocabulary to remain compatible with LCSH, thus improving subject interoperability between Dublin Core and MARC metadata.

Multilingual linking and mapping are used to provide access to multiple languages across systems. The MACS (Multilingual Access to Subjects) research project uses a linking method approach.[114] Three vocabularies (LCSH, RAMEAU,[115] and SWD[116]) are manually matched and linked to each other by using conceptual clusters, thus allowing each subject heading list to remain autonomous. The research team also developed, tested, and refined a Linked Management Interface (LMI) that assists in link creation. Landry notes, however, that since the matching process is done manually, "the basic task of establishing links will remain very time-consuming."[117]

Classification systems are often used as an intermediary or switching language between different languages and different KOS because they are perceived as being language independent. Kwasnik and Rubin studied the impact of the differences in knowledge structures from language to language and culture to culture by examining a set of terms related to the concepts of family, which are universal yet culturally bound.[118] The purpose of the study was to identify problems surfacing from the extension of a source classification system to accommodate another language and culture and to suggest methods to deal with the problem. Their research first explored the differences in kinship terms and relationships through interviews with 14 informants of diverse language and cultural backgrounds, using ethnographic interview techniques that included

creating a visual display of the family relationships. Kwasnik and Rubin then compared the resultant terms to the way in which the *Dewey Decimal Classification* (DDC) and the *Library of Congress Classification* (LCC) expressed kinship terms and relationships. This study revealed seven problematic patterns, and the researchers suggested ways to deal with some of the problems. The area of language and culture is an important one for academic libraries with diverse user populations. This significant study needs replication with multiple informants for each language/culture and a more comprehensive range of languages.

The OCLC Research Office has a strong research agenda for vocabulary mapping, automatic classification, and automatic term assignment. In the late 1990s, OCLC launched the Scorpion project to explore the automatic classification of electronic resources by creating clusters of terms based on statistical and linguistic methodologies.[119] Scorpion software used the DDC data files "to create a searchable database of subject terminology used to describe each Dewey number in the Dewey schedules."[120] Classification categories are mapped to DDC and create predefined subject clusters. The document is then filtered through the predefined clusters as a query and Dewey classes are assigned in ranked order. Scorpion software was integrated into the OCLC CORC project to automatically class internet resources using terms extracted from the document. The researchers view the use of a language-independent notation scheme as a way to enable different language translations of DDC vocabulary and captions to link to a Dewey class number and allow the user to select the appropriate language view.[121] Other research undertaken by the OCLC team includes development of the DeweyBrowser for library collections, mapping different vocabularies to each other to create a linked semantic system that will accommodate cross-domain searching, and the WordSmith Toolkit, which extracts words and phrases from full-text documents.[122] In addition, information about many other current research projects can be found at the OCLC ResearchWorks site (http://www.oclc.org/research/researchworks/) and in the journals *Knowledge Organization*, *Journal of Documentation*, and *Journal of the American Society for Information Science and Technology.*

The Library OPAC

The library OPAC is the focus of the final section of research discussed in this chapter. When all is said and done, the goal of research in knowledge

organization is to improve the quality, functionality and usability of library catalogs and databases for information retrieval. OPAC research consists of two primary areas. One sphere of research is quality assessment and evaluation of existing knowledge organization systems, including data content, record syntax, and system performance. The other sphere is applied research to enhance the functionality and usability of existing systems and to develop new systems. Of necessity the research discussed here is selective and limited to those projects that specifically address library catalogs. For a broader discussion of system evaluation, one should refer to the information science literature.

Many different research methodologies are used in evaluative research, often in combination with one another. Common methods include transaction log and content analysis, grounded theory, survey questionnaires, focus groups and interviews, think-aloud protocols, and observation. Two articles in the 1990s by Seymour[123] and Large and Beheshti[124] provided in-depth reviews of the OPAC research conducted from the mid-1980s to the mid-1990s. Much of the research discussed in these two reviews involves evaluation and quality assessment.

A 1996 special issue of *JASIS* was devoted to OPAC research and contained several articles that focus on various aspects of research on this topic.[125] Following up on an earlier article, Borgman asks the question "Why are Online Catalogs *Still* So Hard to Use?"[126] In her theoretical analysis, the author cites two underlying causes for the problem: the lack of focus on user information-seeking behavior to inform OPAC design and ongoing emulation of the card catalog search process. Hert examined the goals of users searching an OPAC (and other multifile systems) as part of a larger project that focused on users' interactions with the OPAC.[127] Using a grounded theory methodology where the data analysis was iterative and generalizations were formed from the data itself, Hert found that although the users expressed their goals to some extent, it was necessary to identify the critical incidents (i.e., when the user became aware of a gap in knowledge, thus creating an information need) that had shaped the goal to gain a fuller understanding of the search. Hert's research also showed that the goals of the search were not greatly modified during the search process. The findings of this study suggest that OPAC design may be improved "by providing detailed information on the situational elements which influence goals and on the potential constancy of goals on such systems."[128]

For several decades, user studies have documented problems with subject searching in OPACs.[129] Larson conducted a longitudinal study of the University of California's MELVYL online catalog over a six-year period, using a transaction log analysis methodology.[130] His results showed that only 12% of subject searches recovered between 1 and 20 items, thus signifying that subject searches were more likely to fail (no hits or too many hits) than to succeed in a very large university library. Larson's findings also indicated a persistent decline in the percentage of subject searches over the data collection period and a corresponding increase in the percentage of known item, especially title keyword, searches. These results suggest that over time users learned to substitute various forms of known item searches for subject searches in response to the frequent failure of subject searches.

Drabenstott and Weller created an experimental OPAC (ASTUTE) that contained two separate catalogs.[131] Both catalogs used extended subject searching functionality, but only one catalog also used subject search trees to direct the system's selection of searching techniques in response to user queries. The study was conducted in two academic libraries, using both quantitative and qualitative methodologies. While the quantitative analysis based on retrieval precision yielded mixed results, the qualitative data gathered from the users was more conclusive. The overall research results showed that the "subject-access design that featured search trees was more effective in selecting a subject-searching approach that would produce useful information for the subjects users seek than users would select on their own."[132]

Another more recent study by Yu and Young used a transaction log analysis (TLA) methodology in a longitudinal study over a two-year period to identify and define unsuccessful OPAC subject searches.[133] The TLA indicated that a major portion of subject searches would have been more successful if the search had been conducted using other bibliographic tools. The authors suggested that this problem is attributed to the ubiquitous use and familiarity with Web search engines and online bookstores and the subsequent mental models that users bring to OPAC searching. Several suggestions were made to improve OPAC searching based on the functionality of Web-based search applications. Graham used TLA to identify and analyze no-hit subject searches in the college OPAC.[134] After noting specific subject searches for which no records were retrieved but for which appropriate

information resources actually were represented in the catalog, Graham explored two potential methods to ameliorate the problem through the use of authority record cross-references and "pathfinder" records providing brief instructions on search refinement. The author also outlines additional steps needed to determine whether the potential ameliorations make a difference to users' subject searching experiences.

Content-focused research looks at the quality and accuracy of the data in catalog/metadata records in order to reduce errors and improve retrieval. An early study by Ballard described a systematic method for finding and eliminating typographical errors in catalogs and listed the most commonly found misspellings.[135] Another study by Ballard and Lifshin analyzed the misspellings and discovered that repeated misspellings tended to have eight or more letters, at least three syllables, and be more common words rather than esoteric technical terms.[136] Ballard and Grimaldi examined errors in MARC tagging to improve OPAC searching.[137] Bowman conducted several studies that examined data content problems that acted as barriers to retrieval.[138] Beall and Kafadar studied the typographical errors in shared bibliographic records to determine whether corrections were made during the copy cataloging process.[139]

MacEwan and Young developed another method for measuring the quality of both the data content and the overall catalog record.[140] These researchers employed the FRBR mapping of user tasks to data attributes and relationships to measure the quality of catalog records in the British Library. Using the data element weighting method developed for FRBR and a sophisticated statistical process, the authors were able to assign an overall quality score to each catalog record examined.

Improving accessibility to networked digital information and resources is another concern of the OPAC research community. Burke, Germain, and Van Ullen conducted survey research to study the effect that adding surrogates for free Web resources would have on the integrity of the catalog.[141] By examining 567 URLs in the OPACs of 24 Association of Research Libraries (ARL) member libraries, they discovered that the percentage of total catalog and maintenance errors varied from 0.0% in three OPACs to a high of 58.32% in one catalog. The authors also found that these libraries did not commonly use persistent URLs, perhaps contributing to the problem. Other OPAC accessibility research examined the effects of OPAC screen changes on search behavior and success.[142]

The metadata community has now matured to the point where studies of evaluation and quality assessment have entered the literature. Initially, research in metadata evaluation was conceptual and addressed the issues of "What is quality?" and "How do we measure it?" Barton, Currier, and Hey studied metadata projects in two communities of practice: learning object repositories and e-print archives.[143] In each project, the researchers analyzed the metadata to identify defects that demonstrated a need for quality assurance in the assignment of metadata. Bruce and Hillman adopted a different approach to defining metadata quality and used a systematic, domain- and method-independent way to identify metadata quality indicators.[144] Kelly, Closier, and Hiom used theoretical and applied research to develop a quality assessment framework for metadata and a self-assessment tool kit.[145] Moen, Stewart, and McClure identified a list of assessment criteria drawn from the literature and applied qualitative and quantitative content analysis techniques to assess the quality of metadata records in the Government Information Locator Service (GILS).[146]

The research on evaluation and quality assessment of catalogs, cataloging, and metadata is immensely important in academic libraries in light of the rapid growth of digital libraries and metadata use. As academic libraries seek and develop new methods for generating automated and partially automated metadata, the importance of quality assessment and evaluation increases significantly. Many digital library projects allow the object creators or curators to assign metadata to their own objects. This type of distributed metadata creation will need quality assurance measures embedded in the process design.[147]

Library 2.0 and OPAC Design Enhancements

The Library 2.0 movement, with its emphasis on service, interactivity, and personalization, calls into question the limited functionality of library OPACs in terms of capitalizing on recent developments in Web service-oriented applications and architecture. In order to plan for the future, several recent studies have analyzed the technological and social trends that could impact major academic and research libraries. In one study commissioned by the LC, Calhoun explored the future of research library catalogs in the context of today's Web search engines, online bookstores, user expectations, and economic considerations.[148] She derived her research questions from an extensive multidisciplinary literature review, which included marketing

and business perspectives along with those of the library science and information retrieval communities. Calhoun then developed a set of structured interview questions from the literature search and identified interviewees from among library and information science leaders, scholars, technologists, systems developers and implementers, vendors, and library practitioners in order to collect different points of view on the future of catalogs. Her purposive sample consisted of 23 experts in the field.[149] After data analysis was completed, Calhoun made 32 recommendations to LC based on the literature review and expert interviews. Recommendations centered on revitalizing catalogs for internal operations as well as the communities they serve and integrating catalogs with other discovery tools. High on the list of recommendations was rethinking catalog content and services to significantly enhance functionality by using many of the features found in popular Web search engines. This report helps set the research agenda for online catalogs for the short-term future, as many areas of the recommendations will require further investigation before they can be implemented. Some members of the library community took issue with Calhoun's marketing framework and realistic assessment of user expectations;[150] however, two other studies, one by the University of California[151] and another by Indiana University[152] support many of Calhoun's findings and suggestions.

Most college students today are technologically savvy and, as Calhoun pointed out, the mental models that they bring to library Web-based catalogs, databases, and portals are often based on their experiences with Web resources such as Amazon.com, Google, MySpace, or the Internet Movie Database. Influenced by Web sites that provide seamless interactive functionality and access to different databases, both commercial vendors and open-source application developers are rapidly designing new interfaces for library catalogs and portals. Many new products incorporate relevance ranking, faceted navigation, meaningful result clustering, visual representation of results, breadcrumb trails, and federated search tools.[153] Some of these applications include Endeca,[154] AquaBrowser,[155] Evergreen,[156] Encore,[157] and Primo.[158] To date, the body of research is small and consists primarily of case studies and anecdotal articles on implementation; however, as these features become more commonplace, the usability and evaluative research no doubt will grow.

Other Library 2.0 concepts of personalization and interactivity (social tagging, adding reviews, blogging comments, etc.) are beginning

to influence library catalog and portal design. The scarce research that exists tends to focus on public libraries or social tagging Web sites. One usability study, however, did examine an academic portal customization software application, *MyLibrary*. Brantley, Armstrong, and Lewis used categories of common tasks to test "the participants' abilities to customize a personal library web page, understand the resource categories as defined by librarians, and manage the discipline-specific content available in the portal."[159] The results of the study indicated that even experienced computer users struggled with customizing the library portal; however, when users overcame the learning curve, they appreciated the shortcuts offered by the portal. This study also revealed that user misunderstanding of librarian-defined resource categories remains a serious stumbling block to library Web site usability.

One way to help overcome the disconnect between user and library vocabularies brings us to another Library 2.0 concept: user interaction with the catalog through social tagging, reviews, and comments. Again it is public libraries that have taken the lead in implementing these features,[160] but academic institutions are not far behind. Plymouth State University developed a prototype open-source OPAC called Scriblio.[161] Based on the WordPress blog software platform, Scriblio has many Library 2.0 features, including a "recent searches" sidebar; book jacket displays; reviews; holdings data; automatic identification of related items based on author and subject data; and a box to "search inside the book" that uses Amazon's Application Programming Interface (API). In addition, each catalog record allows comments, trackbacks, and tagging. Although these front-end OPAC applications are too new to have generated a body of research, it is expected that this area of investigation will grow rapidly in the near future.

Conclusions

For the past 15 years, researchers have sought to develop ways to automate the systems and processes of knowledge organization. The digital environment presented opportunities for academic libraries to explore alternative methods for organizing digital resources, as seen by the many new metadata schemas that have been developed and the application of XML markup to provide more flexibility in these systems. Another research trend shows a move toward developing systems of organization

that are simpler and easier to apply than the currently used rules, tools, and processes. The approaches of the Dublin Core Metadata Initiative (DCMI) and the FAST project indicate not only a need for simplicity, but also an attempt to extend the use of library organizational practices outside the library community, and in turn to use metadata created by these other communities. The distributed networked environment, coupled with a growing emphasis on global cooperation, is the catalyst for interoperability research that is key to both cataloging and metadata knowledge organization systems.

When the research conducted in the field of knowledge organization and its impact on academic libraries are examined, other trends prevail that represent major paradigm shifts in the approach to organizing information. The reconceptualization of the fundamental principles and structures undergirding the view of the bibliographic universe and the organizational process has led to the adaptation of catalog and record structures that will allow more meaningful displays of multiple versions of a work and will express the relationships among works more clearly; both issues must be dealt with on a daily basis by catalogers in academic libraries with collection depth. This reconceptualization was the catalyst for the ongoing revision and restructuring of the *AACR2* cataloging code that is used by academic libraries. The question remains as to whether the final version of the RDA code will be a contemporary set of rules that is based on the research and trends discussed in this chapter and will appeal to the broadening knowledge organization community, or an outdated rehash of the current system with only a passing nod to the dramatic changes facing academic libraries today.

None of the research projects discussed here exists in a vacuum. They were undertaken to solve problems of a practical nature or refute incorrect assumptions. Many of the projects were carried out in academic environments or with academic library needs as the focal point, since academic and research libraries have the most to gain from automation and applications that can facilitate multilingual subject access and streamline the cataloging, classification, and metadata creation processes. It is appropriate therefore, that academic libraries serve as test beds for research that will ultimately be used or implemented to assist in the task of knowledge organization.

The integration of cataloging and metadata creation is fast becoming the norm in academic libraries and will require an expanded view

of data curation.[162] Knowledge organization specialists will be called upon to work collaboratively with systems and information technology personnel to develop the means of discovery and access for information distributed throughout the university. Catalogers will be increasingly involved in research projects to develop and evaluate digital libraries, electronic dissertations and theses, institutional repositories, and large research data sets.

This research review indicates that the knowledge organization research community is thriving and is responsible for many of the rapid changes taking place in libraries today. In order to remain at the cutting edge of this field, however, more extensive research studies are needed. The expanding diversity of our academic communities demands better search and display capabilities for nonroman script materials. We must identify and develop alternative actions for failed searches and better navigation of large retrieval sets. We have only just begun to investigate the effects of Library 2.0 concepts such as expanded record content (e.g., tables of contents, images, music snippets, etc.) and user interactivity (e.g., assigning tags and adding reviews and comments) on catalog usage and library resources. Additional services such as those provided by online bookstores that use active agents for recommendations, relevance ranking, and improved precision and recall need to be studied. More research is required for catalog-based alert services and customization of the library catalog and portal. There is much to be learned from Internet services, but in order to move forward we must explore the possibility of successful implementation of such services and their impact on the academic community. It is unfortunate that in today's economic climate most research is funded by agencies for short-term projects (one to three years), rather than supporting researchers to develop and implement long-term research agendas. Researchers dedicated to the improvement of knowledge organization and retrieval should not be content with reaching Google or Amazon.com benchmark functionalities, but should look beyond these popular services and begin to identify new avenues of research for the knowledge organization community of the future.

References

1. Elaine Svenonius, "Bibliographic Control," in *Academic Libraries: Research Perspectives,* ed. Mary Jo Lynch and Arthur P. Young, 38–66 (Chicago: American Library Association, 1990).

2. Ibid., 42–43.
3. IFLA UBCIM Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report* (Munich: K.G. Saur, 1998).
4. Eeva Murtomaa,"The Impact of the Functional Requirements for Bibliographic Records Recommendations on the ISBD(ER)" *Cataloging and Classification Quarterly* 28, no. 1 (1999): 34.
5. VTLS was the first vendor to experiment with developing a FRBR structured catalog. Some issues and options for developing the system are presented in a PowerPoint presentation by Vinod Chachra and John Espley, "Differentiating Libraries through Enriched User Searching: FRBR as the Next Dimension in Meaningful Information Retrieval," www.ala.org/ala/alcts/education/presentations/ac04/ac04presentations.cfm (accessed March 30, 2008).
6. Rick Bennett, Brian F. Lavoie, and Edward T. O'Neill, "The Concept of a Work in WorldCat: An Application of FRBR," *Library Collections, Acquisitions, and Technical Services* 27, no. 1 (2003): 45–59.
7. Edward T. O'Neill, "FRBR: Functional Requirements for Bibliographic Records: Application of the Entity-Relationship Model to Humphry Clinker," *Library Resources and Technical Services* 46, no. 4 (2002): 150–159.
8. Roberto Sturman, "Implementing the FRBR Conceptual Approach in the ISIS Software Environment: IFPA (ISIS FRBR Prototype Application)," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 253–70; Stefan Gradmann, "RDFS:FRBR : Towards an Implementation Model for Library Catalogs Using Semantic Web Technology," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 63–75.
9. Thomas B. Hickey, Edward T. O'Neill, and Jenny Toves, "Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR)," *D-Lib Magazine* 8, no. 9 (2002), http://www.dlib.org/dlib/september02/hickey/09hickey.html (accessed March 31, 2008); Thomas B. Hickey and Edward T. O'Neill, "FRBRizing OCLC's Worldcat," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 239–251.
10. Jackie Radebaugh and Corey Keith, "FRBR Display Tool," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 253–270.
11. Chew Chiat Naun, "FRBR Principles Applied to a Local Online Journal Finding Aid," *Library Resources and Technical Services* 51, no. 2 (2007): 134–145.
12. Ketil Albertsen and Carol van Nuys, "Paradigma: FRBR and Digital Documents," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 125–149; Tom Delsey, "Modeling Subject Access: Extending the FRBR and FRANAR Conceptual Models," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 49–61; Kerry Kilner, "The AustLit Gateway and Scholarly Bibliography: A Specialist Implementation of the FRBR," *Cataloging and Classification* 39, no. 3/4 (2005): 87–102.

13. Robyn Holmes and Marie-Louise Ayres, "MusicAustralia: Towards a National Music Information Infrastructure" (paper presented at the International Symposium for Music Information Retrieval [ISMIR] Conference, Barcelona, Spain, October 10–14, 2004), online at http://www.nla.gov.au/nla/staffpaper/2004/ayres2.html (accessed March 30, 2008); Gunilla Jonsson. "Cataloguing of Hand Press Materials and the Concept of Expression in FRBR." *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 77–86; Patrick LeBoeuf., "Musical Works in the FRBR Model or 'Quasi La Stessa Cosa': Variations on a Theme by Umberto Eco," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 103–124; Simon C. Lin and others, "A Metadata Case Study for the FRBR Model Based on Chinese Painting and Calligraphy at the National Palace Museum in Taipei"(paper presented at the International Conference on Dublin Core and Metadata Applications, Tokyo, October 24–26, 2001), online at http://www.nii.ac.jp/dc2001/proceedings/product/paper-08.pdf (accessed March 30, 2008); David Miller and Patrick LeBoeuf, "'Such Stuff as Dreams Are Made On': How Does FRBR Fit Performing Arts?" *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 151–178; Yann Nicolas, "Folklore Requirements for Bibliographic Records: Oral Traditions and FRBR," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 179–195.
14. Glenn Patton, "Extending FRBR to Authorities," Cataloging and Classification Quarterly 39, no. 3/4 (2005): 39–48.
15. Shoichi Taniguchi, "Expression-Level Bibliographic Entity Records: A Trial on Creation from Pre-Existing MARC Records," *Cataloging and Classification Quarterly* 38, no. 2 (2004): 33–59.
16. Shoichi Taniguchi, "Conceptual Modelling of Component Parts of Bibliographic Resources in Cataloguing," *Journal of Documentation* 59, no. 6 (2003): 692–708.
17. Tom Delsey and others, *The Logical Structure of the Anglo-American Cataloguing Rules, Part I* (Joint Steering Committee for Revision of AACR, August 1998): 1, http://www.collectionscanada.ca/jsc/docs/aacr.pdf (accessed March 30, 2008); see also Tom Delsey with Beth Dulabahn and Michael Heaney, *The Logical Structure of the Anglo-American Cataloguing Rules. Part II* (Joint Steering Committee for Revision of AACR, January 1999), http://www.collectionscanada.ca/jsc/docs/aacr2.pdf (accessed March 30, 2008).
18. Barbara B. Tillett, "RDA: Resource Description and Access: Background and Context" (paper presented at the ALA Annual Conference. Chicago, IL, June 6, 2005), http://www.ala.org/ala/alcts/alctsconted/alctsceevents/alctsannual/AACR3_Tillett.pdf (accessed March 30, 2008).
19. Gordon Dunsire, "Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization," *D-Lib Magazine* 13, no. 1/2 (2007), http://www.dlib.org/dlib/january07/dunsire/01dunsire.html (accessed March 30, 2008).
20. Tom Delsey, *Mapping ISBD Elements to FRBR Entity Attributes and Relationships:*

A Report (Den Haag, Netherlands: International Federation for Library Associations, 2004), 1, http://www.ifla.org/VII/s13/pubs/ISBD-FRBR-mappingFinal.pdf (accessed March 30, 2008).

21. Patrick Le Boeuf, "The Impact of the FRBR Model on the Future Revisions of the ISBDs: A Challenge for the IFLA Section on Cataloguing," *International Cataloguing and Bibliographic Control* 31, no. 1 (2002): 3–6.
22. Tom Delsey, "Functional Analysis of the MARC 21 Bibliographic and Holdings Formats," 2002, http://www.loc.gov/marc/marc-functional-analysis/original_frbr.html (accessed March 30, 2008).
23. Library of Congress, Network Development and MARC Standards Office, "Functional Analysis of the MARC 21 Bibliographic and Holdings Formats," 2006, Library of Congress, http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html (accessed March 30, 2008).
24. Barbara B. Tillett, "Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Information Used in Cataloging" (PhD diss., University of California, Los Angeles, 1987).
25. Tillett, "Bibliographic Relationships," 190.
26. Richard P. Smiraglia, "Authority Control and the Extent of Bibliographic Relationships" (PhD diss, University of Chicago, 1992).
27. Sherry L. Vellucci, "Bibliographic Relationships Among Musical Bibliographic Entities: A Conceptual Analysis of Music Represented in a Library Catalog with a Taxonomy of the Relationships Discovered" (DLS diss., School of Library Service, Columbia University, 1995); Sherry L. Vellucci, *Bibliographic Relationships in Music Catalogs* (Lanham, MD: Scarecrow Press, 1997).
28. Sherry L. Vellucci," Bibliographic Relationships," in *The Principles and Future of AACR: Proceedings of the International Conference on the Principles and Future Development of AACR Toronto, Canada, October 23–25, 1997,* ed. Jean Weihs, 105–146 (Toronto: Canadian Library Association; Chicago: American Library Association, 1998).
29. Martha M. Yee, "What Is a Work? Part 1: The User and the Objects of the Catalog," *Cataloging and Classification Quarterly* 19, no. 1 (1994): 33–40; Martha M. Yee, "What Is a Work? Part 2: The Anglo-American Cataloging Codes," *Cataloging and Classification Quarterly* 19, no. 2 (1994): 5–22; Martha M. Yee, "What Is a Work? Part 3: The Anglo-American Cataloging Codes," *Cataloging and Classification Quarterly* 20, no. 1 (1995): 25–46; Martha M. Yee, "What Is a Work? Part 4: Cataloging Theorists and a Definition," *Cataloging and Classification Quarterly* 20, no. 2 (1995): 3–24.
30. Allyson Carlyle, "The Second Objective of the Catalog: An Evaluation of Collocation in Online Catalog Displays" (PhD diss, University of California, Los Angeles, 1994).
31. Allyson Carlyle, "User Categorisation of Works: Toward Improved Organisa-

tion of Online Catalog Displays," *Journal of Documentation* 55, no. 2 (1999): 184–208.

32. Allyson Carlyle, "Ordering Author and Work Records: An Evaluation of Collocation in Online Catalog Displays," *Journal of the American Society for Information Science* 47, no. 7 (1996): 538–554.
33. Allyson Carlyle and Joel Summerlin, "Transforming Catalog Displays: Record Clustering for Works of Fiction," *Cataloging and Classification Quarterly* 33, no. 3/4 (2002): 13–25.
34. Richard P. Smiraglia, *The Nature of "a Work": Implications for the Organization of Knowledge.* (Lanham, MD: Scarecrow Press, 2001).
35. Patrick Le Boeuf, "'Convergence Is the Goal': Activity Report of the IFLA/FRBR/CIDOC CRM Harmonization Group" (PowerPoint presentation made at FRBR in 21st Century Catalogs: An Invitational Workshop, Dublin, OH, May 2–4, 2005), http://www.oclc.org/research/events/FRBR-workshop/presentations/leboeuf/FRBR_CRM_Harmonization.ppt (accessed March 30, 2008).
36. Bennett, Lavoie, and O'Neill, "The Concept of a Work."
37. Richard P. Smiraglia, ed., "Works as Entities for Information Retrieval," special issue, *Cataloging and Classification Quarterly* 33, no. 3/4 (2002).
38. Kristin Antelman, "Identifying the Serial Work as a Bibliographic Entity," *Library Resources and Technical Services* 48, no. 4 (2004): 250.
39. See the discussion papers presented at FRBR in 21st Century Catalogs: An Invitational Workshop, Dublin, OH, May 2–4, 2005, http://www.oclc.org/research/events/FRBR-workshop/program.htm (accessed March 30, 2008).
40. Antelman, "Identifying the Serial Work as a Bibliographic Entity," 240.
41. Jean Hirons and Crystal Graham, "Issues Related to Seriality," in *The Principles and Future Development of AACR: Proceedings of the International Conference on the Principles and Future Development of AACR: Toronto, Ontario, Canada, October 23–25, 1997,* ed. Jean Weihs, 180–212 (Chicago: American Library Association, 1998).
42. Ibid., 183.
43. Arlene G. Taylor, ed., *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools* (Westport, CT: Libraries Unlimited, 2007). This collection of essays includes the following chapters: Arlene G. Taylor, "An Introduction to Functional Requirements for Bibliographic Records (FRBR)," 1–21; Glenn E. Patton, "An Introduction to Functional Requirements for Authority Data (FRAD)," 21–28; Glenn E. Patton. "Understanding the Relationship between FRBR and FRAD," 29–34; William Dent, "FRBR and the History of Cataloging," 35–58; Edward T. O'Neill, "The Impact of Research on the Development of FRBR," 59–72; Richard P. Smiraglia. "Bibliographic Families and Superworks," 73–86; Barbara B. Tillett.. "FRBR and RDA (Resource Description and Access)," 87–96; Alexander C. Thurman. "FRBR and Archival

Materials," 97–102; Murtha Baca and Sherman Clarke, "FRBR and Works of Art, Architecture, and Material Culture," 103–110; Mary Lynette Larsgaard, "FRBR and Cartographic Materials," 111–116; Martha M. Yee, "FRBR and Moving Image Materials," 117–130; Sherry L. Vellucci, "FRBR and Music," 131–152; and Steven C. Shadle, "FRBR and Serials," 153–174.

44. IFLA Meeting of Experts on an International Cataloguing Code, "Statement of International Cataloguing Principles" (April 2006 draft), International Federation of Library Associations and Institutions, http://www.loc.gov/loc/ifla/imeicc/pdf/statement-draft3_apr06cleancopy.pdf (accessed March 30, 2008).
45. IFLA Meeting of Experts on an International Cataloguing Code, "Statement of International Cataloguing Principles," 1.
46. Two entire issues of the *Journal of Internet Cataloging* (vol. 4, nos. 1 and 2, 2001) are devoted to the CORC project.
47. Keith E. Shafer, "Mantis Project: A Toolkit for Cataloging," in *Annual Review of OCLC Research 1998* (published online by OCLC in 1999), archived at http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003483 (accessed March 30, 2008); Keith E. Shafer, "Kilroy: An Internet Research Project," in *Annual Review of OCLC Research 1996* (published online by OCLC in 1997), archived at http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003364 (accessed March 30, 2008); Keith E. Shafer, "Scorpion Helps Catalog the Web," *Bulletin of the American Society for Information Science and Technology* 24, no. 1 (1997): 28–29.
48. Eric Childress, "Crosswalking Metadata in the OCLC CORC Service," 81–88; Carol Jean Godby and Ray Reighart, "Terminology Identification in a Collection of Web Resources," 49–66; Vizine-Goetz, Diane. "Dewey in CORC: Classification in Metadata and Pathfinders," 67–80; Thomas B. Hickey, "Collaboration in CORC," 5–16; all in *Journal of Internet Cataloging* 4, no. 1/2 (2001).
49. Ingrid Hsieh-Yee and Michael Smith, "The CORC Experience: Survey of Founding Libraries, Part 1," *OCLC Systems and Services* 17, no. 3 (2001):133–140; Ingrid Hsieh-Yee and Michael Smith, "The CORC Experience: Survey of Founding Libraries, Part 2," *OCLC Systems and Services* 17, no. 4 (2001):166–177.
50. Tschera Harkness Connell and Chandra Prabha, "Characteristics of Resources Represented in the OCLC CORC Database," *Library Resources and Technical Services* 46, no. 2 (2002): 39–49.
51. Diane Vizine-Goetz, "Dewey in CORC: Classification in Metadata and Pathfinders," *Journal of Internet Cataloging* 4, no. 1/2 (2001): 67–80; Edward T. O'Neill and Brian F. Lavoie, "Bibliographic Control for the Web," *Serials Librarian* 37, no. 3 (2000): 53–69; Lorraine Normore, "CORC Interface Designed for Experienced Catalogers as Well as Other Users," *OCLC Newsletter,* no. 239 (1999): 35.

52. Keith Shafer and others, "Introduction to Persistent Uniform Resource Locators," (paper presented at INET96, Montreal, Quebec, Canada, June 24–28, 1996), http://www.isoc.org/inet96/proceedings/a4/a4%5F1.htm (accessed March 30, 2008).
53. Two journals in particular cover the organization of electronic resources. *Serials Librarian* includes articles on all types of seriality, much of which includes electronic resources, and *The Journal of Internet Cataloging* is devoted exclusively to the organization of all types of electronic resources.
54. Karen S. Calhoun, "Redesign of Library Workflows: Experimental Models for Electronic Resource Description," in *Bicentennial Conference on Bibliographic Control for the New Millennium: Conference Proceedings,* ed. Ann M. Sandberg-Fox, 357–376 (Washington, DC: Library of Congress, Catalog Distribution Center, 2001).
55. Xiaotian Chen and others, "E-Resource Cataloging Practices: A Survey of Academic Libraries and Consortia." *Serials Librarian* 47, no. 1/2 (2004): 153–179.
56. Ibid., 174.
57. Charity K. Martin and Paul S. Hoffman, "Do We Catalog or Not? How Research Libraries Provide Bibliographic Access to Electronic Journals in Aggregated Databases," *Serials Librarian* 43, no. 1 (2002): 61–77.
58. Yiu-On Li and Shirley W. Leung, "Computer Cataloging of Electronic Journals in Unstable Aggregator Databases: The Hong Kong Baptist University Library Experience," *Library Resources and Technical Services* 45, no. 4 (2001): 198–211.
59. David Banush, Martin Kurth, and Jean Pajerek, "Rehabilitating Killer Serials: An Automated Strategy for Maintaining E-Journal Metadata," *Library Resources and Technical Services* 49 no. 3 (July 2005): 190–203.
60. Ann Copeland, "E-Serials Cataloging in the 1990s: A Review of the Literature," *Serials Librarian* 41, no. 3/4 (2002): 7–29.
61. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), "Functional Requirements for Authority Data: A Conceptual Model" (April 1, 2007 draft), International Federation of Library Associations and Institutions, http://www.ifla.org/VII/d4/FRANAR-ConceptualModel-2ndReview.pdf (accessed March 30, 2008).
62. IFLA Working Group on the Functional Requirements for Subject Authority Records, "Terms of Reference," IFLANET, http://www.ifla.org/VII/s29/wgfrsar.htm (accessed March 30, 2008).
63. Arlene G. Taylor, "Variations in Personal Name Access Points in OCLC Bibliographic Records," *Library Resources and Technical Services* 36, no. 2 (1992): 224–241.
64. Tamara S. Weintraub, "Personal Name Variations: Implications for Authority Control in Computerized Catalogs," *Library Resources and Technical Services* 35, no. 2 (1991): 217–228.

65. J. H. Bowman, "The Catalog as Barrier to Retrieval, Part 2: Forms of Names," *Cataloging and Classification Quarterly* 30, no. 4 (2000): 51–73.
66. Qiang Jin, "Comparing and Evaluating Corporate Names in the National Authority File (LCNAF) on OCLC and on the Web," *Cataloging and Classification Quarterly* 36, no. 2 (2003): 21–30.
67. Karen S Calhoun and W. Michael Oskins, "Rates and Types of Changes to LC Authority Files," *Information Technology and Libraries* 1, no. 2 (1992): 132–136.
68. R. Ruiz-Pérez, E. Delgado López-Cózar, and E. Jiménez-Contreras, "Spanish Personal Name Variations in National and International Biomedical Databases: Implications for Information Retrieval and Bibliometric Studies," *Journal of the Medical Library Association (JMLA)* 90, no. 4 (2002): 411–430.
69. Jiajian Hu, "Transactional Analysis: Problems in Cataloging Chinese Names," *Illinois Libraries* 82, no. 4 (2000): 251–260.
70. H.-C. Bolick, "Problems in the Establishment of Nonunique Chinese Personal Headings with Special Reference to NACO Guidelines and Vendor-Supplied Authority Control," *Library Resources and Technical Services* 43, no. 2 (1999): 95–105.
71. Yewang Wang, "A Look into Chinese Persons' Names in Bibliography Practice," *Cataloging and Classification Quarterly* 31, no. 1 (2000): 51–81.
72. MARBI Multilingual Record Task Force, "Multilingual Authority Records in the MARC 21 Authority Format," Discussion Paper 2001-DP05 (2001), Library of Congress, http://www.loc.gov/marc/marbi/2001/2001-dp05.html (accessed March 31, 2008).
73. Martha Speirs Plettner, "Arabic Name Authority in the Online Environment: Options and Implications," *International Cataloguing and Bibliographic Control* 32, no. 2 (2003): 23–26.
74. Deborah E. Kulczak, "Name Authority Work for OCLC Copy Cataloging: Is It Worth the Effort?" *Cataloging and Classification Quarterly* 28, no. 1 (1999): 69–81.
75. Evan Pappas, "An Analysis of Eight RLIN-Members' Authority-Controlled Access Points for Purposes of Speeding Copy Cataloging Work Flow," *Cataloging and Classification Quarterly* 22, no. 1 (1996): 29–47.
76. Jeffrey Beall. "Using OCLC Connexion to Find Typographical Errors in Authority Records," *OCLC Systems and Services* 20, no. 2 (2004): 71–75.
77. Nadine P. Ellero, "Panning for Gold: Utility of the World Wide Web for Metadata and Authority Control in Special Collections," *Library Resources and Technical Services* 46, no. 3 (2002): 79–84, 87–91.
78. Qiang Jin, "Creating Up-to-Date Corporate Name Authority Records by Using Official Corporate Home Web Pages," *Cataloging and Classification Quarterly* 38, no. 3/4 (2004): 281–290; Qiang Jin, "Is the Current Way of Constructing Corporate Authority Records Still Useful?" *Information Technology and Libraries* 24, no. 2 (2005): 68–76.

79. Robert E. Wolverton. "Authority Control in Academic Libraries in the United States: A Survey," *Cataloging and Classification Quarterly* 41, no. 1 (2005): 111–131.
80. N. S. Bangalore and Chandra G. Prabha, "Authority Work in Copy (Derived) Cataloging: A Case Study," *Technical Services Quarterly* 15, no. 4 (1998): 39–56.
81. Nedria A. Santizo and Charlene S. Rezabek, "Series Authority Control: Report of a Survey," *Cataloging and Classification Quarterly* 15, no. 1 (1992): 75–81.
82. Susan Lee Tsui, and Carole F. Hinders, "Cost-Effectiveness and Benefits of Outsourcing Authority Control," *Cataloging and Classification Quarterly* 26, no. 4 (1999): 43–61.
83. Tim DiLauro and others, "Automated Name Authority Control and Enhanced Searching in the Levy Collection {Computer File}," *D-Lib Magazine* 7, no. 4 (2001), http://www.dlib.org/dlib/april01/dilauro/04dilauro.html (accessed March 30, 2008).
84. Mark Patton and others, "Toward a Metadata Generation Framework: A Case Study at Johns Hopkins University," *D-Lib Magazine* 10, no. 11 (2004), http://www.dlib.org/dlib/november04/choudhury/11choudhury.html (accessed March 30, 2008).
85. Ibid.
86. Ibid.
87. Karen E. Greever, "A Comparison of Pre- and Post-Cataloging Authority Control," *Library Resources and Technical Services* 41, no. 1 (1997): 39–49.
88. James C. French and Allison L. Powell, "Using Clustering Strategies for Creating Authority Files," *Journal of the American Society for Information Science* 51, no. 8 (2000): 774–786.
89. David Miller, "XOBIS: An Experimental Schema for Unifying Bibliographic and Authority Records," *Cataloging and Classification Quarterly* 39, no. 3/4 (2005): 285–303.
90. Jutta Weber, "LEAF: Linking and Exploring Authority Files," *Cataloging and Classification Quarterly* 38, no. 3/4 (2004): 227–236.
91. Vinh-The Lam, "Outsourcing Authority Control: Experience of the University of Saskatchewan Libraries," *Cataloging and Classification Quarterly* 32, no. 4 (2001): 53–69.
92. Sherry L. Vellucci, "Commercial Services for Providing Authority Control: Outsourcing the Process," *Cataloging and Classification Quarterly* 39, no. 1/2 (2004): 443–456.
93. Arlene G. Taylor, and others, eds., *Authority Control in Organizing and Accessing Information: Definition and International Experience* (Binghamton, NY: Haworth Information Press, 2004).
94. For an early overview of metadata development see Sherry L. Vellucci, "Metadata," in *Annual Review of Information Science and Technology 33*, ed. Martha

E. Williams 187–222 (Medford, NJ: Information Today for the American Society for Information Science, 1998).

95. Jane Greenberg and others, "Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization," *Journal of Digital Information* 2, no. 2 (2002).
96. Jane Greenberg and others, "Iterative Design of Metadata Creation Tools for Resource Authors" (paper presented at 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice—Metadata Research & Applications, Seattle, WA, September 28–October 2, 2003), http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf (accessed March 31, 2008).
97. Amanda J. Wilson, "Toward Releasing the Metadata Bottleneck: A Baseline Evaluation of Contributor-Supplied Metadata," *Library Resources and Technical Services* 51, no. 1 (2007): 16–28.
98. "Répertoire International de Littérature Musicale," http://www.rilm.org/ (accessed March 30, 2008).
99. DiLauro and others, "Automatic Name Authority Control"; Patton and others, "Toward a Metadata Generation Framework."
100. Ozgur Yilmazel, Christina M. Finneran, and Elizabeth D. Liddy, "MetaExtract: An NLP System to Automatically Assign Metadata," in *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2004, Tucson, AZ, USA, June 7–11, 2004, Proceedings*, ed. Hsinchun Chen and others, 241–242 (ACM, 2004).
101. Gordon W. Paynter, "Developing Practical Automatic Metadata Assignment and Evaluation Tools for Internet Resources," in *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2005, Denver, CO, USA, June 7–11, 2005, Proceedings,* ed. Mary Marlino, Tamara Sumner, and Frank M. Shipman III, 291–300 (ACM, 2005).
102. Ibid., 291.
103. American Library Association, Committee on Cataloging: Description and Access (CC:DA), "Task Force on Metadata: Final Report," American Library Association (2002), http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html (accessed March 31, 2008).
104. National Information Standards Organization, *Understanding Metadata* (Bethesda, MD: NISO Press, 2004), 13, http://www.niso.org/publications/press/UnderstandingMetadata.pdf (accessed March 31, 2008).
105. Carol Jean Godby, Jeffrey Young, and Eric Childress, "A Repository of Metadata Crosswalks," *D-Lib Magazine* 10, no. 12 (2004), http://www.dlib.org/dlib/december04/godby/12godby.html (accessed March 31, 2008).
106. Ibid.
107. Ibid.
108. Carl Lagoze and Herbert Van de Sompel, "The Making of the Open Archives Initiative Protocol for Metadata Harvesting," *Library Hi-Tech* 21, no. 2 (2003):

118–128.

109. Herbert Van de Sompel and others, "Resource Harvesting within the OAI-PMH Framework," *D-Lib Magazine* 10, no. 12 (2004), http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html (accessed March 31, 2008).
110. Tina Gross and Arlene G. Taylor, "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results," *College and Research Libraries* 66, no. 3 (2005): 215.
111. Ibid.
112. Marcia Zeng and Lois Mai Chan, "Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems," *Journal of the American Society of Information Science and Technology* 55, no. 5 (2003): 377–395.
113. Lois Mai Chan and Edward T. O'Neill, "FAST (Faceted Application of Subject Terminology): A Simplified Vocabulary Based on the Library of Congress Subject Headings," *IFLA Journal* 29, no. 4 (2003): 336–342.
114. Patrice Landry, "Multilingual Subject Access: The Linking Approach of MACS," *Cataloging and Classification Quarterly* 37, no. 3/4 (2004): 177–191.
115. "Recherche Autorités RAMEAU," Catalogue Bn-Opale Plus, http://catalogue.bnf.fr/jsp/recherche_autorites_rameau.jsp?host=catalogue (accessed March 31, 2008)
116. *SWD (Schlagwortnormdatei)* refers to the subject heading list. *RSWK (die Regeln für den Schlagwortkatalog)* refers to the rules for creating a subject heading list. Both are discussed in detail in Konrad Umlauf, *Einführung in die Regeln für den Schlagwortkatalog RSWK,* Berliner Handreichungen zur Bibliothekswissenschaft., v. 66 (Berlin: Institut für Bibliothekswissenschaft der Humboldt-Universität zu Berlin, 1999), http://www.ib.hu-berlin.de/~kumlau/handreichungen/h66/ (accessed March 31, 2008)
117. Landry, "Multilingual Subject Access," 187.
118. Barbara H. Kwasnik and Victoria L. Rubin, "Stretching Conceptual Structures in Classifications across Languages and Cultures," *Cataloging and Classification Quarterly* 37, no. 1/2 (2004): 33–47.
119. Srividhya Subramanian and Keith Shafer, "Clustering," *Journal of Library Administration* 34, no. 3/4 (2001): 221–228.
120. Thomas B. Hickey and Diane Vizine-Goetz, "The Role of Classification in CORC," *Journal of Library Administration* 34, no. 3/4 (2001): 425.
121. Ibid., 424–425.
122. Online Computer Library Center, "Learn More about the DeweyBrowser," OCLC, http://www.oclc.org/research/researchworks/ddc/browser.htm (accessed March 31, 2008); Diane Vizine-Goetz and others, "Vocabulary Mapping for Terminology Services," *Journal of Digital Information* 4, no. 4 (2004), http://journals.tdl.org/jodi/article/view/jodi-128/113 (accessed March 31, 2008); C. Jean Godby and Ray R. Reighart, "The WordSmith

Toolkit," *Journal of Library Administration* 34, 3/4 (2001): 307–316.

123. Sharon Seymour, "Online Public-Access Catalog User Studies: A Review of Research Methodologies, March 1986–November 1989," *Library and Information Science Research* 13, no. 2 (1991): 89–102.
124. Andrew Large and Jamshid Beheshti, "OPACs: A Research Review," *Library and Information Science Research* 19, no. 2 (1997): 2, 111–133.
125. Micheline Beaulieu and Christine L. Borgman, eds., "Current Research in Online Public Access Systems," special issue, *Journal of the American Society for Information Science* 47, no. 7 (1996).
126. Christine L. Borgman, "Why Are Online Catalogs Still Hard to Use?" *Journal of the American Society of Information Science* 47, no. 7 (1996): 493–503.
127. Carol A. Hert, "User Goals on an Online Public Access Catalog," *Journal of the American Society for Information Science* 47, no. 7 (1996): 504–518.
128. Ibid., 516.
129. Seymour, "Online Public Access Catalog User Studies"; Jon R. Hufford, "Use Studies and OPACS," *Technical Services Quarterly* 9, no. 1 (1991): 57–70.
130. Ray Larson, "The Decline of Subject Searching: Long-Term Trends and Patterns of Index Use in an Online Catalog," *Journal of the American Society for Information Science* 42, no. 3 (1991): 197–215.
131. Karen M. Drabenstott and Marjorie S. Weller, "Failure Analysis of Subject Searches in a Test of a New Design for Subject Access to Online Catalogs," *Journal of the American Society for Information Science* 47, no. 7 (1996): 519–537.
132. Ibid., 519.
133. Holly Yu and Margo Young, "The Impact of Web Search Engines on Subject Searching in OPAC," *Information Technology and Libraries* 23, no. 4 (2004):168–180.
134. Rumi Y. Graham, "Subject No-Hits Searches in an Academic Library Online Catalog: An Exploration of Two Potential Ameliorations," *College and Research Libraries* 65, no. 1 (2004): 36–54.
135. Terry Ballard, "Spelling and Typographical Errors in Library Databases," *Computers in Libraries* 12, no. 6 (1992): 14–17.
136. Terry Ballard and Arthur Lifshin, "Prediction of OPAC Spelling Errors through a Keyword Inventory," *Information Technology and Libraries* 11, no. 2 (1992): 139–145.
137. Terry Ballard and Anthony Grimaldi, "Improve OPAC Searching by Reducing Tagging Errors in MARC Records," *Library Software Review* 16, no. 3 (1997): 152–155.
138. J. H. Bowman, "The Catalog as Barrier to Retrieval, Part 1: Hyphens and Ampersands in Titles," *Cataloging and Classification Quarterly* 29, no. 4 (2000): 39–60; Bowman, "The Catalog as Barrier, Part 2"; J. H. Bowman, "Sic Catalog Syndrome: Title Page Transcription as a Barrier to Retrieval," *Cataloging and*

Classification Quarterly 32, no. 1 (2001): 39–54.

139. Jeffrey Beall and Karen Kafadar, "The Effectiveness of Copy Cataloging at Eliminating Typographical Errors in Shared Bibliographic Records," *Library Resources and Technical Services* 48, no. 2 (2004): 92–101.
140. Andrew MacEwan and Thurstan Young, "Quality vs. Quantity: Developing a Systematic Approach to a Perennial Problem," *Catalogue and Index,* no. 152 (2004): 1–7.
141. Gerald Burke, Carol Anne Germain, and Mary K. Van Ullen, "URLs in the OPAC: Integrating or Disintegrating Research Libraries Catalogs," *Journal of Academic Librarianship* 29, no. 5 (2003): 290–297.
142. Deborah D. Blecic, Josephine Dorsch, and Melissa Koenig, "A Longitudinal Study of the Effects of OPAC Screen Changes on Searching Behavior and Searcher Success," *College and Research Libraries* 60, no. 6 (1999): 515–530.
143. Deborah D Barton, Sarah Currier, and Jessie M. N. Hey, "Building Quality Assurance into Metadata Creation: An Analysis Based on the Learning Objects and E-Prints Communities of Practice" (paper presented at 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice—Metadata Research & Applications, Seattle, WA, September28–October 2, 2003), http://www.siderean.com/dc2003/201_paper60.pdf (accessed March 31, 2008).
144. Thomas R. Bruce and Diane I. Hillman, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in *Metadata in Practice,* ed. Diane I. Hillman and Elaine L. Westbrooks, 238–256 (Chicago: American Library Association, 2004).
145. Brian Kelly, Amanda Closier, and Debra Hiom, "Gateway Standardization: A Quality Assurance Framework for Metadata," *Library Trends* 53, no. 4 (2005): 637–650.
146. William E Moen, Erin L. Stewart, and Charles R. McClure, "The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an Exploratory Study" (paper presented at Second IEEE Metadata Conference, Washington, DC, September 15–16 1997, http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm. (accessed March 31, 2008).
147. See the New Jersey Digital Highway, a new project to digitize special collections related to New Jersey, which requires librarians and curators to create the metadata for their collection: http://www.njdigitalhighway.org/librarians.php (accessed March 31, 2008).
148. Karen Calhoun, "The Changing Nature of the Catalog and its Integration with Other Discovery Tools: A Report Prepared for the Library of Congress," Cornell University, February 2006, http://www.loc.gov/catdir/calhoun-report-final.pdf (accessed March 31, 2008).
149. Ibid., Appendix A, 23–27.
150. Thomas Mann, ""The Changing Nature of the Catalog and Its Integration

with Other Discovery Tools. Final Report. March 17, 2006. Prepared for the Library of Congress by Karen Calhoun': A Critical Review," Library of Congress Professional Guild, AFSME Local 2910 (April 3, 2006), http://www.guild2910.org/AFSCMECalhounReviewREV.pdf (accessed March 31, 2008).

151. University of California, Bibliographic Services Task Force, "Rethinking How We Provide Bibliographic Services for the University of California," final report (December 2005), University of California Libraries, http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf (accessed March 31, 2008).
152. Jackie Byrd and others, "A White Paper on the Future of Cataloging at Indiana University," (January 15, 2006), Indiana University, http://www.iub.edu/~libtserv/pub/Future_of_Cataloging_White_Paper.doc (accessed March 31, 2008).
153. Marshall Breeding, "Thinking About Your Next OPAC," *Computers in Libraries* 27, no 4 (2007): 28–30.
154. The Endeca Web site is at http://endeca.com/ (accessed March 31, 2008). To view an academic library's implementation of Endeca, see the catalog of North Carolina State University at http://www.lib.ncsu.edu/catalog/ (accessed March 31, 2008).
155. The AquaBrowser Web site is at http://www.aquabrowser.com/. To view an academic library's implementation of AquaBrowser, see Trinity College, Dublin's Stella Catalogue at http://stella.tcd.ie/ (accessed March 31, 2008).
156. The Evergreen ILS Web site is at http://gapines.org/opac/en-US/skin/default/xml/index.xml (accessed March 31, 2008).
157. Information on Encore is on the Innovative Interfaces Web site at http://www.iii.com/products/encore.shtml. To view an academic library's implementation of Encore, see Yale Law Library at http://morris.law.yale.edu/ (accessed March 31, 2008).
158. Information on Primo is on the ExLibris Web site at http://exlibrisgroup.com/category/PrimoOverview. To view an academic library's implementation of Primo, see Vanderbilt University's Acorn Catalog at https://acorn.library.vanderbilt.edu/ (accessed March 31, 2008).
159. Steve Brantley, Annie Armstrong, and Krystal M. Lewis, "Usability Testing of a Customizable Library Web Portal," *College and Research Libraries* 67, no. 2 (2006): 146.
160. See the Ann Arbor District Library catalog for an example of user-added tags: http://www.aadl.org/catalog (accessed March 31, 2008).
161. The Scriblio Web site it at. http://about.scriblio.net/about. See the Lamson Library implementation of Scriblio at http://library.plymouth.edu/read/223702. (accessed March 31, 2008).
162. Andrew Treloar, David Groenwegen, and Catherine Harboe-Ree, "The Data

Curation Continuum: Managing Data Objects in Institutional Repositories," *D-Lib Magazine* 13, no. 9/10 (2007), http://www.dlib.org/dlib/september07/treloar/09treloar.html (accessed March 31, 2008).