Serials and Multiple Versions, or the Inexorable Trend toward Work-Level Displays
Julian Everett Allgood

Cross-lingual Name and Subject Access
Jung-ran Park

Searching Titles with Initial Articles in Library Catalogs
Clément Arsenault and Elaine Ménard

Electronic Resources Communications Management
Celeste Feather

Automated Access Level Cataloging for Internet Resources at Columbia University Libraries
Kate Harcourt, Melanie Wacker, and Iris Wolley
2 Essential Cataloging & Classification Tools on the Web

**CATALOGER’S DESKTOP**

The most widely used cataloging documentation resources in an integrated, online system—accessible anywhere.

- Look up a rule in AACR2 and then quickly and easily consult the rule’s LC Rule Interpretation (LCRI).
- Turn to dozens of cataloging publications and metadata resource links plus the complete MARC 21 documentation.
- Find what you need quickly with the enhanced, simplified user interface.

**Free trial accounts & annual subscription prices:**

Visit [www.loc.gov/cds/desktop](http://www.loc.gov/cds/desktop)

For free trial, complete the order form at [www.loc.gov/cds/desktop/OrderForm.html](http://www.loc.gov/cds/desktop/OrderForm.html)

AACR2 is the joint property of the American Library Association, the Canadian Library Association, the Chartered Institute of Library and Information Professionals.

---

**CLASSIFICATION WEB**

Full-text display of all LC classification schedules & subject headings. Updated daily.

- Find LC/Dewey correlations—Match LC classification and subject headings to Dewey® classification numbers as found in LC cataloging records. Use in conjunction with OCLC’s WebDewey® service for perfect accuracy.
- Search and navigate across all LC classes or the complete LC subject headings.

**Free trial accounts & annual subscription prices:**

Visit [www.loc.gov/cds/classweb](http://www.loc.gov/cds/classweb)

For free trial, complete the order form at [www.loc.gov/cds/classweb/application.html](http://www.loc.gov/cds/classweb/application.html)

Dewey and WebDewey are registered trademarks of OCLC Online Computer Library Center, Inc.
Library Resources & Technical Services (ISSN 0024-2527) is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Association for Library Collections & Technical Services, a division of the American Library Association. Subscription price: to members of the Association for Library Collections & Technical Services, $27.50 per year, included in the membership dues; to nonmembers, $75 per year in U.S., Canada, and Mexico, and $85 per year in other foreign countries. Single copies, $25. Periodical postage paid at Chicago, IL, and at additional mailing offices. POSTMASTER: Send address changes to Library Resources & Technical Services, 50 E. Huron St., Chicago, IL 60611. Business Manager: Charles Wilt, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Send manuscripts to the Editorial Office: Peggy Johnson, Editor, Library Resources & Technical Services, University of Minnesota Libraries, 490 Wilson Library, 309 19th Ave. So., Minneapolis, MN 55455; (612) 624-2312, fax: (612) 626-9353, e-mail: m-john@umn.edu. Advertising: ACLTS, 50 E. Huron St., Chicago, IL 60611; 312-280-5038, fax: 312-280-5032. ALA Production Services: Troy D. Linker, Angela Hanshaw, Angela Gwizdala, and Chris Keech. Members: Address changes and inquiries should be sent to Membership Department—Library Resources & Technical Services, 50 E. Huron St., Chicago, IL 60611. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to Library Resources & Technical Services, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 1-800-545-2433; fax: (312) 944-2641; subscription@ala.org.

Library Resources & Technical Services is indexed in Library Literature, Library & Information Science Abstracts, Current Index to Journals in Education, Science Citation Index, and Information Science Abstracts. Contents are listed in CALL (Current American—Library Literature). Its reviews are included in Book Review Digest, Book Review Index, and Review of Reviews.

Instructions for authors appear on the Library Resources & Technical Services Web page at www.ala.org/alcts/lrts. Copies of books for review should be addressed to Edward Swanson, Book Review Editor, Library Resources & Technical Services, 1065 Portland Ave., Saint Paul, MN 55104, e-mail: swans152@umn.edu.

©2007 American Library Association

All materials in this journal subject to copyright by the American Library Association may be photocopied for the noncommercial purpose of scientific or educational advancement granted by Sections 107 and 108 of the Copyright Revision Act of 1976. For other reprinting, photocopying, or translating, address requests to the ALA Office of Rights and Permissions, 50 E. Huron St., Chicago, IL 60611.


Library Resources & Technical Services does not imply official endorsement by the Association for Library Collections & Technical Services nor by ALA, and the assumption of editorial responsibility is not to be construed as endorsement of the opinions expressed by the editor or individual contributors.

ISSN 0024-2527  July 2007  Volume 51, No. 3

Editorial  158

ARTICLES

Serials and Multiple Versions, or the Inexorable Trend toward Work-Level Displays  160
Julian Everett Allgood

Cross-lingual Name and Subject Access: Mechanisms and Challenge  180
Jung-ran Park

Searching Titles with Initial Articles in Library Catalogs: A Case Study and Search Behavior Analysis  190
Clément Arsenault and Elaine Ménard

NOTES ON OPERATIONS

Electronic Resources Communications Management  204
A Strategy for Success  206
Celeste Feather

Automated Access Level Cataloging for Internet Resources at Columbia University Libraries  212
Kate Harcourt, Melanie Wacker, and Iris Wolley

Book Reviews  226

Index to Advertisers  228

ABOUT THE COVER

Catalog Department, Los Angeles Public Library, October 15, 1948. West side of room showing typists and catalogers at work. Photo courtesy ALA Archives, Urbana, Illinois.

Visit LRTS online at www.ala.org/alcts/lrts. For current news and reports on ALCTS activities, see the ALCTS Newsletter Online at www.ala.org/alcts/alcts_news.
Editorial

Peggy Johnson

This issue of LRTS includes papers that address critical issues in technical services associated with description, discovery, and access—and related issues about managing electronic resources. Julian Everett Allgood, in “Serials and Multiple Versions, or the Inexorable Trend toward Work-Level Displays,” takes on the complexities of serials description and the problems of multiple versions in his thoughtful exploration of how we got to where we are and what we need to do to improve access for users. How to manage e-resources is the topic of two papers that examine solutions developed in two libraries. Kate Harcourt, Melanie Wacker, and Iris Wolley describe “Automated Access Level Cataloging for Internet Resources at Columbia University Libraries.” The often thorny problem of keeping current with the many messages needed to manage electronic resources is considered by Celeste Feather in her paper, “Electronic Resources Communications Management: A Strategy for Success.” Feather explains the use of a communications audit to analyze the types of communication that move through a technical services unit.

Two papers look at the confusion we can create for catalog users. Jung-ran Park contemplates “Cross-lingual Name and Subject Access: Mechanisms and Challenge,” alerting us to cultural and linguistic problems that can limit access, using the Korean language as an example. Clément Arsenault and Elaine Ménard report on how users search in “Searching Titles with Initial Articles in Library Catalogs: A Case Study and Search Behavior Analysis.” They analyze the confusion and problems current systems can cause and propose alternatives.

LRTS continues to change to meet the needs of our readers. We are advancing toward our goal of making LRTS a journal for the twenty-first century. One of our innovations is to send ALCTS members an “early alert,” which offers brief introductions to the contents. These alerts soon will contain live links to the articles. We are making one article in each issue available now through the Web site (www.ala.org/alcst/lrts). To start the year and volume 51, no. 1, we posted John J. Riemer’s guest editorial, “Restrategizing Bibliographic Services and the One Good Record.” “Mapping WorldCat’s Digital Landscape,” by Brian F. Lavoie, Lynn Silipigni Connaway, and Edward T. O’Neill (from no. 2) is available now.

I am delighted to announce a new policy that grants current and past LRTS authors the rights to post their articles to an institutional or disciplinary repository. No additional permission is required. The official statement is:

The American Library Association grants authors whose works are published in the Library Resources & Technical Services journal (“LRTS Journal”) permission to submit a PDF of the author’s published paper (the “Work”) to an institutional or disciplinary repository, so long as the Work is not modified or altered and the author cites the LRTS Journal, including volume, issue, and date, as place of original publication. Author may not make any other publication of the Work without ALA’s prior written express consent. This policy applies to both past and future articles published in the LRTS Journal.
Is managing your e-journal collection more difficult than you expected?

We can help.

From obtaining an accurate list of the titles you’ve ordered to handling registration issues to troubleshooting access problems and more, managing electronic collections can require more time and attention than you have to give.

EBSCO’s services include e-journal audits to confirm that your library is billed only for the titles ordered, itemized invoices to facilitate budget allocation and customized serials management reports to assist with collection development.

Our dedicated e-journal customer service team assists with non-access problems, IP address changes and more. And our suite of e-resource access and management tools minimizes administrative tasks while maximizing patron experience.

Let us put our expertise to work for you. Contact your EBSCO sales representative today.

www.ebsco.com
The proliferation of multiple versions for bibliographic works presents numerous challenges to the cataloger and, by extension, to the catalog user. Fifteen years after the Multiple Versions Forum held in Airlie, Virginia, online public access catalog (OPAC) users continue to grapple with confusing displays representing numerous serial manifestations (i.e., versions) resulting from the Anglo-American Cataloguing Rules' (AACR2) cardinal principle (Rule 0.24). Two initiatives offer hope for more coherent OPAC displays in light of a renewed focus upon user needs: the ongoing revision of AACR2, and the International Federation of Library Associations and Institutions’ Functional Requirements for Bibliographic Records (FRBR) model. A third potential tool for improving OPAC displays exists within a series of standards that have developed to parallel library needs, and today offer a robust communications medium: the MARC 21 authority, bibliographic, and holdings formats. This paper summarizes the challenges posed by multiple versions and presents an analysis of current and emerging solutions.

A dilemma confronts the Anglo-American cataloging community. Library catalogs display multiple occurrences of titles available in different formats as multiple hits for a user’s search query, rather than clustering them into a single entry or hit. The variety of formats and versions of resources libraries collect continues to grow, yet the underlying manifestation level principles of the Anglo-American Cataloguing Rules, 2nd ed. (AACR2) result in catalogs difficult for users to navigate.¹ This multiple versions (MulVer) problem represents a defining challenge of the automated catalog era.

This paper will examine the MulVer problem with regard to serial resources and will consider both the Joint Steering Committee for Revision of Anglo-American Cataloguing Rules (JSC) mandate to revise AACR2 and the growing influence of the Functional Requirements for Bibliographic Records (FRBR) model.² As my conclusions are aimed at current and developing solutions to the MulVer problem, the literature I cite was written largely within the last fifteen years. The paper calls for online public access catalog (OPAC) displays allowing users to more easily understand and navigate the rich, complex collections librarians assemble.

Unless otherwise noted, the term “users” refers to external library users rather than to library staff members. For library staff members or internal library users, manifestation-level detail is necessary for ordering, record identification,
check-in, and other library functions. That a given data element serves a purpose for internal library staff, however, does not necessitate its display to all library users. Much of the manifestation-level detail of AACR2 serial bibliographic records currently displayed in library OPACs is inconsequential for external library users. Users are more interested in obtaining the journal article content than in the manifestation-level details of the serial title in which the article is published. As indicated by Lubetzky, researchers typically approach the library OPAC with a citation to a specific issue of a specific volume of a specific serial work. They simply need to know if the collection contains the serial title and issue containing the selected article. Library catalogs following AACR2 Rule 0.24 contain a separate OPAC record for each version or manifestation of each serial work or expression. For serial titles that many library catalogs contain in multiple physical formats, these separate OPAC records for equivalent versions further increase the likelihood for user confusion.

Antelman has illustrated that the core responsibility of librarians and library catalogs remains to guide users to the content they seek. In the case of serial resources, users seek content at the article level more often than at the title or physical manifestation level. Thus the first obstacle users must overcome in order to identify, select, and obtain serial resources within library catalogs is that librarians long ago abdicated the role of providing article-level journal citations to abstracting and indexing agencies. Library catalogs typically provide title-level access to their journal collections. It is left to others to provide users with citations to the wealth of content within each of these serial titles. In addition to providing access to journal titles, library catalogs have historically done an admirable job of informing users of the various formats or versions serials are issued in, and the means for using them. For example, the full serial run of The New Yorker on CD-ROM would be of little use without access to a computer able to display the disc contents. Today's users prefer that everything be available online, but they still routinely use articles on paper or microform. Again, users simply want to know if the library has the journal content (i.e., article) they need. They are confused and frustrated by library catalogs forcing them to examine separate records for each format or manifestation. Based on how users struggle with serial multiple versions, today's librarians and library catalogs are not fulfilling the core responsibility of guiding users to content.

**Rule 0.24 and Manifestation-Level Cataloging in AACR2**

AACR2, the *International Standard Bibliographic Descriptions (ISBDs)* and the *International Standard Serial Number (ISSN)* Manual are presently undergoing significant revision with emphasis upon addressing user needs. This therefore seems an ideal time to reconsider some of the underlying precepts and principles of cataloging.

The cardinal principle of AACR2 Rule 0.24 is the foundation for manifestation-level cataloging, which results in record displays that confuse and frustrate users. From AACR2's initial publication in 1978 until 2002, Rule 0.24 read (with minor wording changes):

> It is a cardinal principle of the use of part I that the description of a physical item should be based in the first instance on the chapter dealing with the class of materials to which that item belongs. . . . In short, the starting point for description is the physical form of the item in hand, not the original or any previous form in which the work has been published.

Many believe this focus upon the physical carrier expressed in Rule 0.24 has resulted in the MulVer problem. For example, Graham concludes, “the logical extension of this cardinal principle is the prescription to create a unique record for almost every variant manifestation of a work.” Others are convinced that nowhere in AACR2, neither in Rule 0.24 nor anywhere else, does the code mandate that catalogers build a separate descriptive record. Attig has written, “This rule does not tell [catalogers] whether or not they must describe each manifestation.”

The 2002 AACR2 rule revision significantly changed Rule 0.24 for the first time. This revision was in response to the Committee on Cataloging: Description and Access (CC:DA) Task Force on Rule 0.24 recommendations, and as Beacom points out, represents a “solid improvement.” The current Rule 0.24 reads:

> It is important to bring out all aspects of the item being described, including its content, its carrier, its type of publication, its bibliographic relationships, and whether it is published or unpublished. In any given area of the description, all relevant aspects should be described. As a rule of thumb, the cataloger should follow the more specific rules applying to the item being cataloged, whenever they differ from general rules.

Despite revision, it is difficult not to read this rule as an instruction to continue cataloging physical carriers. The phrase “item being described/item being cataloged” appears twice, and “carrier information” is second in the list of enumerated attributes. Carrier information is not unimportant; yet according to user studies it is not the most important manifestation-level attribute to be described.
Manifestation-level cataloging has become the default norm within AACR2 cataloging for two reasons. The first centers upon cooperative cataloging. In a cooperative cataloging environment in which individual libraries exchange surrogate descriptions to facilitate and enhance user access to their collections of manifestations, describing those shared surrogates at the manifestation level is logical. Within today’s shared cataloging environment in which millions of records are available to libraries through bibliographic utilities such as OCLC and RLIN, it is critical that cataloging and acquisitions librarians be able to select specific manifestations to import into their library catalogs. Describing resources at the manifestation level enables library personnel to do so. The second reason for using manifestation-level records is that libraries need bibliographic records to serve duties beyond their most visible role as surrogate descriptions within the OPAC. Many of the administrative functions librarians perform such as ordering, check-in, and claiming are manifestation specific. Librarians need today’s integrated library management systems (ILMS) to utilize records serving both purposes. Software designers developing ILMS systems must understand this duality of purpose.

Departures from Manifestation-Level Cataloging within Current AACR2 Practice

As Howarth points out, “While the cataloguing code is explicit in its directives for handling different manifestations of the same title or work, application of those rules has been less than consistent.”12 For example, Chapter 11 of AACR2 describes manifestation-level cataloging for microform resources. However, a Library of Congress Rule Interpretation (LCRI) for Chapter 11 instructs catalogers to base their descriptions on the original resource rather than the microform in hand.13 Most American libraries follow the rule interpretation rather than AACR2. Those libraries that follow the LCRI in effect clone the manifestation level record for the original, and add a note describing the microform holdings. For those libraries with holdings of both the original resource and the microform reproduction, as is often the case with serial resources, this creates a cluster of virtually identical, separate records for the title that users must then view one at a time. This is frustrating. This deviation from strict manifestation-based cataloging results in confusing records within OPACs and conflicting records within the internationally shared bibliographic utilities.

Two current Cooperative Online Serials Program (CONSER) practices also deviate from manifestation-level cataloging in favor of a more pragmatic, user-oriented approach. The first is CONSER’s single-record approach.14 During the 1990s, many serials catalogers balked at the prospect of adding yet another bibliographic record for yet another equivalent online version to their local OPACs to remain in accord with national policy and AACR2 Rule 0.24. In response, CONSER developed an alternate approach, allowing catalogers to append descriptive and access attributes for online manifestations to existing print descriptions. In theory, this represented a clear, practical solution to a pressing problem. This technique of providing access to two separate manifestations upon a single bibliographic description led to worries about how ILMS systems would continue the double duty of OPAC display and administrative functionality. Having responded to earlier requests to recognize and handle distributed, consortial library structures and to adhere to the MARC 21 Holdings standard, ILMS systems provided a technique for libraries to attach multiple holdings records along with the individual check-in and receiving attributes necessary to coordinate these separate manifestations or versions. ILMS software designers had therefore cleared a significant hurdle of the MultiVer problem.

The practical implications for libraries willing to extrapolate from CONSER’s single-record guidelines were immense. Citing the precedent set by guidelines of the United States Newspaper Project, and in response to a clear user preference and need, some libraries began to bundle all equivalent serial manifestations upon a single bibliographic description.15 This requires selecting one manifestation to serve as a serial work description or springboard with all equivalent manifestations attached as a holdings record. While attaching and displaying multiple manifestations to a single bibliographic description within some ILMS systems is both possible and practical, sharing or exchanging manifestation and holdings information across our cooperative, distributed cataloging environment is difficult.

The second current CONSER practice that strays from manifestation-level cataloging is the aggregator-neutral record.16 Approved in 2003, aggregator-neutral records reflect the reality that not only are more serial titles available online, many of these online journals are simultaneously available from more than one provider or aggregator. The aggregator-neutral record allows catalogers to create a single bibliographic description representing an online serial and then attach as many access paths or URLs as necessary. When providers subsequently add or remove titles from their packages of electronic journals, catalogers simply add or remove the corresponding URL rather than having to create or delete entire bibliographic descriptions. Figure 1 is an example of a CONSER aggregator-neutral serial record. This particular title is available online from Project Muse, JSTOR, and Ingenta, among others.

Despite the benefit to users, sharing these records within a cooperative cataloging environment is difficult. By providing access to multiple serial manifestations, these bibliographic records come dangerously close to compromising the integrity of the MARC 21 standard as applied within the AACR2 environment. In the single-record approach, descriptions of print and online manifestations of a serial include an $56$ field (used for electronic access and location
Serials and Multiple Versions

At the 2005 CONSER Operations Committee Meeting, it was determined that including the 007 field in records using the single-record technique causes confusion for the ISSN Centers and other user communities. Therefore, CONSER will write and present a discussion paper to the American Library Association’s (ALA) Machine-Readable Bibliographic Information (MARBI) Committee proposing a one-byte “electronic online resource” value for the 008/23 byte. Confusion for some user communities arises because the majority of the record describes the print manifestation. Only by reading and understanding a 530 note (additional physical formats available) detailing the availability of an online version will users comprehend why a record describing the 007 and 856 fields are included upon what otherwise appears to be a print description. Furthermore, this bundling of multiple serial manifestations on a single bibliographic record complicates the batch processing capabilities of automated systems.

These two CONSER practices are admirable in attempting to provide a means of displaying equivalent serial versions to facilitate the needs of users. Within today’s MARC 21 and AACR2 environment, these two CONSER practices create problems for users and the automated systems upon which libraries rely. Librarians and ILMS systems designers need to consider user preferences in providing access to serial resources. If librarians decide to modify the descriptive preferences and access guidelines for serial resources within the revised cataloging code and also modify the MARC 21 communications formats libraries use for exchanging records, the immediate results may include enhanced record sharing and display capabilities. When the JSC circulated the AACR3 draft of Part 1 for comments in early 2005, one prominent concern raised in the ALA response was that the draft failed to address either the MulVer problem or the single-record approach many libraries use to minimize its effects.

Just as AACR2 Rule 0.24 is sometimes interpreted as not mandating manifestation-level cataloging, it may similarly be read as not requiring cohesive manifestation-level displays. The important principle within Rule 0.24 is that catalogers portray specific manifestation-level attributes. Only through doing so can catalogs and OPACs achieve Cutter’s third objective of describing for users all available editions/versions/manifestations of a work. How these manifestation-level attributes are best communicated and displayed to users through the MARC 21 authority, bibliographic, and holdings formats is a decision best left to catalogers and catalog designers. Having demonstrated that departures from manifestation-level cataloging exist today, we need to look more closely at the MulVer problem.

The MulVer Dilemma—Development and Recognition of the Multiple Versions Problem

By the early 1980s, libraries recognized what is now known as the Multiple Versions, or MulVer, problem (also sometimes referred to as the format variation problem). In 1989, Graham wrote a seminal paper addressing the reasons for its emergence and identifying the problems MulVer has wrought upon catalogs. Some argue that the MulVer problem stems primarily from strict adherence to the cardinal principle of AACR2 (Rule 0.24). Additional factors have contributed, as well. Mandel indicates that technological advances within the publishing industry and especially electronic publishing, coupled with the preservation reproductions commissioned by libraries, have contributed to numerous versions of many works. In today’s era of digital manifestations, the MulVer problem has only increased. Weiss states, “Since electronic data can be republished at almost no cost, multiple versions,
The MulVer problem persisted throughout the 1990s, and in 1999 CC:DA assembled the Task Force on Rule 0.24, which revised Rule 0.24 to lessen its emphasis upon the physical carrier. In so doing, the Task Force was certainly aware of the precedent set by the recently revised and republished International Standard Bibliographic Description, Electronic Resources (ISBD [ER]) of 1997. Therein, Weiss writes that for the first time, an international standard allows:

The inclusion of all physical forms of the content on the same bibliographic record [thereby enabling] the record to focus on the content of the work. The physical forms of the work become subordinate instances of the intellectual work, which clearly shows the influence of research done on bibliographic relationships by Barbara Tillett and others (including the International Federation of Library Associations and Institutions (IFLA) Study Group on the Functional Requirements of the Bibliographic Record). In this case, works that have what Tillett refers to as “equivalence relationships,” e.g., works where the authorship and intellectual content are identical, were grouped together on a single record. Conceptually, this was a shift from AACR2 1988 (with its emphasis on specific item description) to the notion that the physical carrier of the information was of only incidental interest to users, who first and foremost would want access to information in whatever form it was available [emphasis added].

What exact role AACR2 Rule 0.24 may eventually play in the new cataloging code, Resource Description and Access (RDA) is unknown. The CC:DA Task Force on Rule 0.24 Final Report has called upon the JSC to add an introductory chapter to the cataloging code that specifically will address a number of big picture topics, including the format variation or MulVer issue. In response, the JSC assembled the Format Variation Working Group (FVWG), and charged it with exploring expression-level cataloging. After realizing how few definable and transcribable attributes exist for the expression level, the FVWG group shifted focus to expression-level collocation, or bringing together all disparate manifestations of a particular expression within a catalog. This led to an exploration of uniform title authority records as a means of distinguishing specific works and expressions within catalogs and of collocating manifestations of the same work and expression. The emphasis moved from the records themselves to the display of the records. This transition itself represents a FRBR influence, as the group went from focusing upon the minutiae of individual catalog records to considering the larger issue of catalogs and displays.
Based on the abilities of today’s library catalogs, the continuing development of the MARC 21 communications standards, further advances in technology and computing, and increasingly sophisticated users, libraries need to reexamine the MulVer problem and the AACR2 principle from which it arises. First, the automation environment in which libraries and information professionals operate today is almost completely different from what it was at the time of the Multiple Versions Forum recommendations in 1989. Today most large libraries have fully automated their processing and have migrated to a second-generation ILMS available from one of only a handful of library automation vendors. This consolidated automation environment facilitates not only the recognition of new functionality and usage models such as FRBR, but also the implementation of innovations considered beneficial to the shared mission and cooperative efforts of libraries around the world.

Second, as the international library automation marketplace has consolidated, libraries and users have benefited from developing standards, harmonizing efforts, and cooperative cataloging. Today, the MARC 21 Format for Holdings Data has matured into a robust carrier sufficiently supportive of significant descriptive and encoded information. The development of the MARC 21 standards and a general movement away from local processing eccentricities has provided cost efficiencies for library budgets. Meanwhile users have benefited from harmonized OPAC result displays. Each of these initiatives has been furthered by enhanced cooperative cataloging efforts.

Third, the Internet and wireless technology have fundamentally transformed the manner in which users access information and conduct research. Howarth has demonstrated that continuing development since the 1989 Multiple Versions Forum has resulted in a generation of catalogs capable of displaying individual records featuring dynamic linking fields able to link across records and across databases.

Finally, users today have no patience for confusing OPAC displays with multiple hits for equivalent resources. Antelman points out, “In order to make our bibliographic data valuable to scholars and others who seek [serial] works, asserting bibliographic control over a higher level of abstraction than has been our practice is necessary.” Marcum of the Library of Congress goes further in admitting, “the detailed attention that we have been paying to descriptive cataloging may no longer be justified.” Howarth and others see a need for bibliographic records or displays that present all manifestations of a work, making the carriers of the manifestations secondary. Resolving the MulVer problem is in libraries’ vital interest as we endeavor to redefine the Anglo-American Cataloguing Rules for a new generation of users.

Resolving Multiple Versions

In confronting the MulVer problem today, librarians have two viable options: change cataloging practices or improve OPAC displays. Yee has recently argued that many of the problems multiple versions present for users could be resolved if catalogers “were allowed [by the cataloging code] to use the MARC 21 holdings format to attach more than one manifestation to a single bibliographic record.” Such an OPAC could then be optimized by providing a “well-designed holdings display [allowing users to sort] holdings by format, by location, by reproduction date, and so on.” Efforts to revise AACR2 are currently underway with a new cataloging code for the Anglo-American community expected in 2009. Within the current cooperative cataloging environment, a cataloging code advocating anything other than manifestation-level descriptions appears unlikely. With millions of existing manifestation-level descriptions populating our catalogs and with a great deal of internal library functionality dependent upon specific manifestations, libraries need to continue to create and have access to manifestation-level descriptions.

This brings us to our second option. OPAC displays have developed far too little since libraries began automating their card catalogs during the 1960s. In spite of today’s hyperlinked, graphics-oriented, Web-based environment, most library OPACs continue to display descriptions as distinct records, little more than an electronic card catalog. Recent offerings such as hot-linked fields and operational URLs appear paltry compared to the technological wizardry available today. ILMS systems designers and developers need to acknowledge that though library systems need to store and exchange data elements as discrete, cohesive units, OPACs are not compelled to display them as such. Coyle indicates, “Using the appropriate data structures, programs can derive a variety of displays and discovery elements from a single [MARC 21] field.” Data storage and data display are two separate and distinct issues easily confused. For example, Attig has indicated that instead of confronting the critical problem of how to display multiple versions within automated catalogs, the Multiple Versions Forum participants presented a resolution for encoding and storing data about multiple versions. Confusing these two issues has represented a major stumbling block in developing pragmatic library database and display designs. Beacom states explicitly, “there are other ways to split and lump” the double-duty bibliographic records librarians need. RDA could instruct catalogers to create manifestation level descriptions, but well-designed OPACs could then generate displays of all equivalent versions, as well as related works and expressions. Beacom believes the development of such capabilities within library OPACs is quite likely during
the next ten years. Improving OPAC display capabilities holds the greater promise for helping librarians resolve the MulVer problem. Two specific initiatives, the FRBR conceptual model and the MARC 21 communications formats, may bring us even closer to this goal.

The Potential of FRBR

In 1998, the International Federation of Library Associations and Institutions’ Section on Cataloging published the Functional Requirements for Bibliographic Records (FRBR). This document notes, “The study has two primary objectives. The first is to provide a clearly defined, structured framework for relating the data that are recorded in bibliographic records to the needs of users of those records. The second objective is to recommend a basic level of functionality for records created by national bibliographic agencies.”

FRBR is not a draft standard, nor is it intended to replace AACR2 or any other cataloging code. FRBR is a systematic, international examination of automated catalogs and the records that comprise them. The study takes the form of a conceptual model and focuses upon three areas:

1. Bibliographic entities and the attributes necessary to describe and access them as well as to distinguish them unambiguously;
2. Relationships between and among bibliographic entities and the relationships bibliographic descriptions share with other external entities such as people, corporate bodies, and subjects; and
3. How users navigate among bibliographic records to find, identify, select, and obtain bibliographic resources within a national bibliography or a library catalog.

The first two focal points allow the model to establish recommendations for “a basic level of functionality for records created by national bibliographic agencies.”

To date, the bulk of intellectual effort on the part of library constituencies worldwide has been upon the first FRBR area, bibliographic entities and their attributes. Of these, the Group 1 entities (work, expression, manifestation, and item) have received by far the most attention. Despite this disproportionate interest in the FRBR lexicon and specifically the Group 1 entities, the FRBR model holds promise in two additional areas. First, FRBR is a conceptual model intended to help librarians consider the catalog more broadly, i.e., how individual records and the relationships among them contribute to the utility of the overall catalog. In essence, the FRBR model encourages librarians to think about catalogs rather than individual records. The second area of promise within FRBR now being more widely recognized is a renewed emphasis upon users and their needs.

Tillett and Smiraglia’s work on bibliographic relationships will play a vital role in database design as libraries and ILMS systems implement FRBR-aware catalogs. Most librarians envision FRBR-aware catalogs based on these underlying relationship structures to be far easier and more intuitive for users to navigate and interpret.

FRBR and AACR2

Serials catalogers have been slow to familiarize themselves with FRBR and with how the model may benefit OPAC displays for serials and continuing resources. Antelman has illustrated many of the complexities associated with defining serial works and with developing serial identifiers adequate to address the needs of the library community, publishers, and abstracting and indexing services. FRBR’s conceptual model is not a perfect match for current AACR2/CONSER serials cataloging, and further studies are needed to clarify some remaining uncertainties. How to define a serial work remains chief among the FRBR decisions needed from the AACR community. The decision is complicated by the fact that our library catalogs commonly contain serial bibliographic records described using several distinct cataloging conventions.

As we consider how the FRBR model may assist librarians and catalog designers improve OPAC displays for serial resources, another important consideration is how each of the FRBR Group 1 entities applies to serials. For many serial works, there is only one work, one expression, and one manifestation, but the potential for many, many items. For these serials, a FRBR-aware OPAC display does not differ significantly from a traditional OPAC display, and the MulVer problem is negligible. Other serials offer multiple manifestations in a range of language and regional editions. For serial works, each of these separate language and regional editions represents a separate expression, but the FRBR edition attribute is troublesome. For monographs and most other library resources, edition statements represent FRBR manifestation-level attributes. In the case of serials, the edition statement is sometimes an expression attribute, sometimes a manifestation attribute. Many serial expressions use what appear to be edition statements to represent numbering attributes (e.g., 2003 ed., and so on). Serial edition statements in this form represent manifestation attributes. Yet when a serial edition statement targets a specific audience (e.g., teacher’s edition), a geographic region (e.g., Northeastern edition), or a language edition, the edition statement represents an expression-level attribute. These serial titles, available in multiple FRBR expressions and multiple physical formats, will benefit most from FRBR and MulVer-aware OPAC displays.

Within a FRBR-aware catalog, work and expression entities will exist only in what are today considered authority files. Exactly what form these serial work and expression
identifiers will take remains an issue very much in debate. Antelman believes that "the [serial] work identifier should be a dumb number, unrelated to existing identifiers associated with the bibliographic entities that it describes, such as titles, [uniform titles], or ISSN..." In contrast, many, including the JSC's FVWG, believe identifiers should be eye-readable uniform titles. One concern about our ability to uniformly assign serial work and expression identifiers is that currently many parallel pre-AACR2 and Successive Entry serial descriptions populate the CONSER database and local library catalogs. The choice of primary access point (i.e., citation) as well as the valid title variants upon serial records entered according to these two cataloging guidelines are different. Consequently, catalogers describing a new serial manifestation within an AACR2 environment for which there is an existing pre-AACR2 record for an equivalent manifestation are confronted with two unpleasant choices: either redescribe a functional pre-AACR2 record as Successive Entry to synchronize the two descriptions, or face the probability that the two records for these equivalent versions will have different primary access points and, therefore, different citations. From a FRBR perspective, two different primary access points represent two different works. The prospect of adopting a cataloging code requiring serial work and expression identifiers understandably gives serialists pause. Serials catalogers wonder if they will be required to create and accept multiple parallel serial work and expression identifiers if libraries continue to allow both pre-AACR2 and Successive Entry serial cataloging descriptions as valid components of library catalogs and bibliographic utilities. That is, for those serial expressions for which pre-AACR2 descriptions exist for one or more manifestations and Successive Entry descriptions exist for other manifestations, will serials catalogers be expected to create parallel work and expression identifiers for both primary access points when they differ?

What to do with these pre-AACR2 records is a complex problem because, like the MulVer problem, it crosses the boundary between AACR and MARC, and also extends from the bibliographic utilities into our local ILMS systems. The issue is further complicated by the fact that from a pragmatic point of view, these pre-AACR2 records remain functional. Because of significant differences between pre-AACR2 and Successive Entry rules for determining choice of entry, it may be advisable for the AACR/CONSER serials community to stop recognizing the validity of coexisting pre-AACR2 and Successive Entry serial descriptions. One prominent example of how pre-AACR2 and Successive Entry serial records differ is that most pre-AACR2 records do not contain uniform titles. Successive Entry serial records commonly contain a uniform title. As uniform titles affect how serial manifestations are cited and the form of their primary access points, parallel pre-AACR2 and Successive Entry descriptions often result in catalog records for equivalent serial versions with different primary access points. Redescribing or recataloging these pre-AACR2 records as Successive Entry would allow serialists to synchronize the primary access points for all equivalent serial manifestations, thereby collocating each version of a serial work or expression. A policy change of this magnitude would be difficult. Arguing for redescribing serial records that function quite well at present is counterintuitive. That said, there is a strong impetus within the current RDA enterprise recommending that an authority records exist for each serial work and expression. Momentum for this directive was furthered by the distribution draft for worldwide comment of the Functional Requirements for Authority Records (FRAR) conceptual model. With this in mind, one reasonable incentive for redescribing (i.e., recataloging) functional pre-AACR2 records may be that following revision, these presently functional records will operate even more efficiently far into the future. A CONSER Task Group on Non-AACR2 Records has been assembled to consider this and other concerns related to pre-AACR2 serial descriptions.

Also, with the upcoming publication of RDA scheduled for 2009, some catalogers may fear that shortly after redescribing all pre-AACR2 serial descriptions as Successive Entry, they will face a similar maintenance initiative when RDA is published. Though understandable, this argument against more consistent serial descriptions in our catalogs and utilities is flawed. During the serial rule revision process from 1998 through 2002, which followed the 1997 International Conference on the Principles and Future Development of AACR (commonly known as the Toronto Conference), several serial entry guidelines, including a return to Latest Entry cataloging, were considered, and Successive Entry serials cataloging was retained. It therefore appears unlikely that the cataloging rules for serials entry will change markedly (if at all) between the AACR2 2002 revision and the initial iteration of RDA. Nonetheless, guidelines for establishing FRBR and FRAR work and expression identifiers for serial resources, with specific regard to the pre-AACR2 and Successive Entry cataloging guidelines, merits further study.

FRBR and Serials

Within the FRBR model, work and expression records contain only such universal attributes as a title or uniform title identifier, subject tracings, and other access points applicable to all manifestations. As FRBR-aware catalogs develop, the manifestation records linked to serial work and expression records will contain more specific descriptive information than the holdings records in today's catalogs. These records may include descriptive information and such identifier elements as ISSN and ISBN. ILMS systems...
will need to develop algorithms capable of searching across multiple levels of work/expression and manifestation entities as demonstrated by Minno’s hierarchical catalog project.\textsuperscript{57} That is, FRBR-aware catalogs must index and retrieve elements or attributes present in both the authority file (i.e., works and expressions) and in the bibliographic/holdings file where manifestation and item data resides. The final report of a recent CC:DA Task Force for the Review of IFLA’s “Guidelines for OPAC Displays” recommends that ILMS systems generate result displays drawn from data within both the bibliographic and authority files.\textsuperscript{59} These result screens would aid user navigation while the dynamic linking capabilities of today’s Web-based OPACs would reduce the number of redundant searches currently required of library catalog users.

ILMS systems also must be able to limit or refine search results based on data elements or attributes at each of these levels. As Yee says in compiling her 2004 MARC 21 shopping list, “put coded information currently in [the leader], 006, 007 and 008 fields in MARC 21 bibliographic and holdings records in the best possible place to allow ready access to both librarians and the public for direct searching of all kinds of categories for dates, language, country of origin, and physical format . . .”\textsuperscript{50} This capability will empower those users who want to see only the online or print resources a library has available.

Much of the data necessary to generate FRBR-aware displays is encoded in MARC 21 catalog records. Bowen, chair of the FVWG, has stated that unique work and expression headings may not be constructed for every resource.\textsuperscript{60} Therefore, catalogers need to consider and suggest additional ILMS systems techniques of collocating and distinguishing works and expressions based on bibliographic and authority data in current library records. Unfortunately, the data within bibliographic records is not always as pristine or rich as librarians might wish. Bowen continues, “Another important lesson learned [by the FVWG] is that the success of projects to FRBRize existing MARC records depends upon the quality of the data [in those records].”\textsuperscript{61} One area that will have a direct impact on creating FRBR and MulVer displays is uniform title assignment. In exploring expression-level collocation, the JSC’s FVWG demonstrated that uniform titles have tremendous potential as descriptive cataloging tags able to both collocate and distinguish related groups of works and expressions. Uniform titles for serials, though, are an AACR2\textsuperscript{2} innovation. Most pre-AACR2 serial descriptions do not contain uniform titles and even within AACR2, assigning uniform titles remains optional for libraries. For those resource descriptions containing uniform titles, there are errant headings and incorrectly assigned headings. Such errors, requiring human review, will be costly to correct. (For example, see the discussion later in this paper concerning figures 2 and 3.)

Librarians need to help ILMS systems developers understand that in asking for FRBR-aware displays and MulVer-aware displays, we are asking for two distinct development lines. Creating a FRBR-aware OPAC display will not resolve the MulVer problem. As Jones has noted, FRBR-aware OPACs will cluster related works, expressions, and manifestations more clearly, but will not free users of the need to consult multiple records for equivalent versions.\textsuperscript{52} FRBR-aware serial displays may display serial works available in multiple expressions and manifestations as a single entry within a headings list (see table 1). Users interested in selecting from among the available expressions of the New York Times or related works within a catalog could select an entry to expand this tree structure (see tables 1 and 2). They may then identify one of the available manifestations by expanding the tree structure yet again (see table 3). The resulting manifestation-level headings in turn may be expandable in cases where the microform manifestation may be available in microfiche and microfilm, and the electronic manifestation may be available as a CD-ROM, diskette, and online. For most works in library catalogs, FRBR-aware search results will be far less voluminous than this particular example. As of December 2001, an analysis of the OCLC WorldCat database projected that almost 80 percent of the approximately 32 million works available were represented by a single manifestation, and would therefore require no further FRBR-aware display modifications.\textsuperscript{61}

An additional element ILMS software designers must bear in mind in order to limit redundant displays is the concept of attribute inheritance detailed in the FRBR model and further described by Coyle and Minno.\textsuperscript{63} Coyle rightly insists that FRBR-based “identifiers allow the creation of functional records at any [entity] level as long as the rules of inheritance are obeyed, such that any lower level [entity] always inherits data elements from the level above it within its functional group.”\textsuperscript{65} FRBR-aware ILMS systems cognizant of the model’s rules of inheritance will allow multi-tier records to generate clear, non-repetitive OPAC displays. This will contribute significantly toward creating OPACs that users are able to navigate and understand easily. Meanwhile MulVer-aware OPAC displays will require a different development effort as described following.

**FRBR and Multiple Versions**

Upon publication, FRBR generated considerable excitement within the library community. Many believed this fresh model would lead to a satisfactory resolution of the MulVer problem. After all, FRBR focuses largely upon relationships within catalogs and, as defined within Tillet’s taxonomy, what closer relationship could two distinct bibliographic resources share than being equivalent versions?\textsuperscript{66}
In 1997, when the FRBR document was still in draft form, Jones wrote an important paper reconsidering the MulVer problem in light of the FRBR model. Jones describes the MulVer problem for serials and concludes with the belief that the AACR community is moving with due deliberation toward the eventual goal of work level cataloging. In 2003, Coyle assessed the impact of FRBR on current development directions within the cataloging and library systems landscape as moving us toward work-level descriptions, or what she termed the “multi-level, multi-functional library systems record.”

Numerous paths could lead to a work-level approach in cataloging. The cataloging community could revise AACR to advocate work-level descriptions, but as demonstrated above such a change would likely come at the expense of both critical current administrative functionality and the legacy manifestation-level data making up today’s catalogs. A somewhat less radical approach might take advantage of the technological capabilities of a well-programmed ILMS able to process existing manifestation records in response to a user’s query and generate both a FRBR and MulVer-aware OPAC display.

In working with ILMS systems developers to create MulVer-aware OPACs, librarians must remind them of the distinct issues of data storage and data display. Libraries have compelling reasons to continue creating and storing bibliographic descriptions at the manifestation level, but these storage packets have nothing to do with how OPACs then display these data packets. For serial resources, a valuable display sequence would allow users to expand work tree structures to the expression level as described above. Upon selecting a particular serial expression, instead of retrieving multiple manifestation entries as in table 3, a MulVer-aware OPAC would assemble each of the manifestation attributes embodying a specific expression (e.g., the daily edition of the New York Times) and display them to the user as a single manifestation-neutral bibliographic description.

Table 1. FRBR-aware OPAC display for the New York Times work entity with an additional expanded view

<table>
<thead>
<tr>
<th>Search results for “All = New York Times”</th>
</tr>
</thead>
<tbody>
<tr>
<td>w₁  New-York Thomsonian</td>
</tr>
<tr>
<td>w₂  New York thrash</td>
</tr>
<tr>
<td>w₃  New York through the eyes of John Sloan and John Marin</td>
</tr>
<tr>
<td>w₄  New York times</td>
</tr>
<tr>
<td>w₅  New York times 60-minute gourmet</td>
</tr>
<tr>
<td>w₆  New York times, 1851–1951: a centenary address</td>
</tr>
<tr>
<td>w₇  New York times Advertising Department series</td>
</tr>
</tbody>
</table>

*When a user selects w₄ for the New York Times, the following expanded display opens*

w₄ New York times

1. Editions of the New York times +
2. Works about the New York times +
3. Works by the New York times +
4. Works related to the New York times +

Note: The + sign indicates that a particular entry may be expanded.

Table 2. FRBR-aware OPAC display for expressions of the New York Times work entity

<table>
<thead>
<tr>
<th>w₁ New York times</th>
</tr>
</thead>
<tbody>
<tr>
<td>e₁ Audio expression(s) +</td>
</tr>
<tr>
<td>e₂ Daily expression(s) +</td>
</tr>
<tr>
<td>e₃ Large-print expression(s)</td>
</tr>
<tr>
<td>e₄ Weekly expression(s) +</td>
</tr>
</tbody>
</table>

Note: The + sign indicates that a particular entry may be expanded.

Table 3. FRBR-aware OPAC display for multiple manifestations (i.e., versions) of one expression of the New York Times work entity

<table>
<thead>
<tr>
<th>w₁ New York times</th>
</tr>
</thead>
<tbody>
<tr>
<td>m₁ Electronic manifestation(s) +</td>
</tr>
<tr>
<td>m₂ Microform manifestation(s) +</td>
</tr>
<tr>
<td>m₃ Print manifestation(s) +</td>
</tr>
</tbody>
</table>

Note: The + sign indicates that a particular entry may be expanded.
This expandable tree-structure entry for serial works within FRBR-aware and MulVer-aware OPAC displays would represent a significant improvement over the multiple hits serial searches often retrieve in today's OPACs. This tree-like display for works with multiple expressions or manifestations represents one of the most intriguing potential features of the FRBR model for library OPACs. With time and development, ILMS systems should soon be able to offer pre- or post-search features allowing users to identify and select the specific resource of interest. Yee has implemented a catalog of moving image materials similar to the one envisioned previously. According to an e-mail message announcing the availability of the UCLA Film and Television Archive catalog, Yee and her staff have cataloged moving image materials at the expression level and then attached multiple MARC 21 holdings records representing physical format variations as well as other slight manifestation-level differences. The Film and Television Archive at UCLA captures manifestation-level title variations by building work-level authority records with extensive cross-references.

To further illustrate the feasibility for such innovative OPAC display technology, consider the results of a cooperative project between the California Digital Library (CDL), the State University of New York (SUNY) system, and Ex Libris. Individual libraries within these two consortia retain manifestation-level records for titles within their local OPACs. For users of the MELVYL (CDL) and SunCat (SUNY) union catalogs, separate manifestation-level records are consolidated through Ex Libris to display a single work or expression-level record detailing the holding institutions and the separate manifestations each holds.

With OCLC's FictionFinder and Curioser projects, librarians are seeing the first commercial and research applications of the FRBR model to catalogs of existing records. At least one major ILMS vendor currently offers a FRBR OPAC. VTLS's Virtua offers libraries the option of implementing the expandable/collapsible FRBR displays discussed in this paper. Other ILMS vendors have FRBR applications in development. Unfortunately, FRBR implementations thus far include only relatively small subsets of the available bibliographic universe of records, and none of the production versions of these products contain any serial works or expressions.

The Promise of MARC 21

FRBR is not the only option for libraries intent on improving today's OPAC displays in response to user needs. For all of its potential, any significant and widespread implementation of FRBR precepts into cataloging codes and integrated library systems remains years away. Meanwhile, other ways of adopting work and expression-level displays in library OPACs offer potential improvements. The MARC 21 authority, bibliographic, and holdings formats provide one alternative. The MARC 21 authority format represents one possible medium for communicating and exchanging work and expression identifiers. Work and expression identifiers are critical for colocating manifestation-level descriptions, descriptions that multiply to create the MulVer problem. The current Library of Congress Action Plan contains the following near-term goal as one of the recommendations suggested at the November 2000 Bicentennial Conference on Bibliographic Control in the New Millennium: "Develop [the] functional requirements to enable the interchange of manifestation records that support internal [i.e., ILMS OPAC] configurations for FRBR [IFLA Functional Requirements for Bibliographic Records] displays for multiple versions; determine supportive cataloging practices; determine any needed MARC 21 enhancements; communicate these to the vendor community." Yee's article addressing FRBR-aware displays offers specific guidance to ILMS software designers for assembling work and expression identifiers from existing MARC 21 data elements in the bibliographic and authority records in today's library catalogs. Yee's proposed OPAC displays for these identifiers bear little resemblance to how these MARC 21 fields and subfields are stored and exchanged, providing further evidence of the important distinction between data storage and data display.

Some complexities are inherent to developing serial identifiers, and varying interpretations remain regarding how these identifiers should be formulated. For the purposes of this paper, presume that the FVWG uniform title approach is selected. Frequent overlap among serial bibliographic and authority records describing the same work or expression occurs, notably with regard to monographic series, which are by definition also serials. Many of these titles have a serial record in the bibliographic file and a corresponding series authority record (SAR) in the Library of Congress, Name Authority File (LC/NAF). In theory, the citation/primary access point on these two records should match, but for reasons previously cited and having to do mostly with the current acceptance of several contradictory serial entry guidelines, this is not always the case. For example, see figures 2 and 3, representing an LC/NAF series authority record (figure 2) and a CONSER bibliographic record for the same work (figure 3). The qualifiers in the uniform title headings do not match. In FRBR terms then, these two headings intended to cite a single work represent two separate works. Such inconsistencies in work and expression headings foster confusion for both internal and external library users. While this poor heading construction is not a direct result of the MulVer problem, it certainly represents one indirect consequence. Within
enormous bibliographic utilities and catalogs where uniform title entries are required to collocate and distinguish numerous serial works and expressions, the potential for inconsistent heading assignment and construction increases. Inconsistent uniform titles then fail to fully collocate the multiple serial expressions within our intricately constructed catalogs. The resulting failure of library OPACs to clearly fulfill both the collocating and distinguishing roles required of uniform title work and expression identifiers leads to user confusion. As libraries move toward work and expression-level OPAC displays, these inconsistent work and expression headings must be corrected. As with the uniform titles previously discussed, many of these inconsistencies between bibliographic uniform titles and series authority records may require human review. Looking ahead, the IFLA Functional Requirements for Authority Records (FRAR) document currently being drafted must emphasize the importance of keeping such headings in accord.

If libraries choose to resolve the MulVer problem by pursuing a multi-tier approach, creating work and expression-level OPAC displays with manifestation-specific details appended within holdings records, each of the MARC 21 formats will require further development. As Eversberg has indicated in response to the FRBR model, “If it comes to a work-oriented approach, the whole dichotomy of bibliographic vs. authority records [must] be re-evaluated.” Referring to the potential of her proposed multifunctional record, Coyle says, “It is reasonable to assume that a future cataloging structure will embody some degree of hierarchy, especially in the need to express the relationships between multiple versions of the same work.” Two serial-specific issues currently under consideration will be critical if libraries pursue this path toward resolving the MulVer problem: the existence of multiple serial entry guidelines and
proposals to approve multiple 1XX fields within authority records.

Though most libraries have followed Successive Entry cataloging since 1981, Latest Entry and pre-AACR2 serial records continue to reside with Successive Entry records in library catalogs and bibliographic utilities. Pre-AACR2 serial records and their effects upon creating serial work and expression identifiers were discussed previously. In Latest Entry cataloging, the latest known title is entered in the 245 field (Title statement). Previous titles are in 247 fields (Former title). One could reasonably argue that even though these title entries exist at the bibliographic as opposed to the authority level, Latest Entry records currently serve as serial work identifiers in documenting and indexing each known title change for entire serial runs. Whether a user searches the current title or a former title within a catalog containing Latest Entry records, a properly indexed ILMS will retrieve the requested record. As such, the AACR community could make the policy decision that these Latest Entry records will remain intact, and that catalogers will not create authority work or expression identifiers for the titles they represent. A cataloging community policy decision of this sort would have little or no impact on MARC 21 format development.

A second important serial-related issue regards the ramifications of authenticating multiple 1XX fields within series authority records. When trying to reconceptualize library catalogs as user-friendly interfaces, one of the fundamental flaws with Successive Entry serials cataloging is the ability to link and display only to the immediately preceding and succeeding titles. The resulting displays make it difficult for library users to navigate among the manifestation records representing a serial run. Yee characterizes this limitation as a series of precarious stepping stones—if any of the titles, or stones, is missing within the catalog being searched, the serial run cannot be assembled.79

This practice is paralleled within name authority records (including series headings). The records contain one 1XX field and, potentially, one preceding 5XX entry and one succeeding 5XX entry. If the MARBI Committee were to approve series authority records containing multiple 1XX fields, catalogers could represent entire serial runs upon a single work/expression record. This redefinition of the series authority record would eliminate the need to delineate earlier and later titles through the 5XX stepping-stone mechanism, and would simultaneously decrease the number of authority records required to represent serial title runs. By linking the appropriate authorized 1XX field to each bibliographic manifestation of a serial work or expression, all would be clustered and displayed for selection by the user. OPAC users searching the serial title from article citations would retrieve a single work/expression entry displaying all linked manifestations available within the catalog, greatly facilitating navigation through complex serial displays. Whether these serial manifestations are described in MARC 21 bibliographic or holdings records is another area requiring further study. The format could develop to support either, but one scenario, presented by Tillett at a 2005 IFLA FRBR Review Group Workshop, completely removes the bibliographic entity from the catalog.80 Works and expressions are formulated through MARC 21 authority records. Manifestation and item information is represented through the MARC 21 holdings format, and these holdings records are then attached directly to authority records.81

This scenario has generated interest because it would provide a more clearly defined communications standard for the attributes common to serials. Serial bibliographic records in today’s OPACs contain an array of data elements representing FRBR work, expression, and manifestation attributes. Describing serial work and expression attributes in authority records in a central, shared catalog such as LC/NAL would allow individual libraries to attach their specific manifestation and item information in locally maintained but universally-accessible (i.e., viewable) holdings records. As Tillett says:

If we had a clear way of identifying the attributes for a particular work/expression/manifestation/item combination, we could theoretically [present] all such combinations for the same work in a single record, and display [only] the needed elements as the application or user specified. There are many ways this could work.82

The MulVer problem could be resolved with the MARC 21 authority, bibliographic, and holdings formats. By authenticating multiple 1XX fields in series authority records, the format also could help resolve the cumbersome display and navigational shortcomings of today’s AACR2 Successive Entry serials record displays. In order to optimize such a proposal, further development would be required in at least three areas:

1. MARC 21 format development to provide greater flexibility in how libraries distribute bibliographic attributes among authority, bibliographic, and holdings record structures;
2. ILMS and systems development to facilitate the indexing and display of data elements across MARC 21 structures; and
3. Utility (i.e., OCLC) and ILMS development to allow libraries to exchange complex, multi-tier records.

Frustrated with the lack of concerted development initiatives on the part of both the library community and ILMS vendors, some libraries have adopted practices and policies enhancing OPAC displays and addressing user needs within
their local catalogs. For instance, the UCLA Film and Television Archives creates expression-level records for moving image materials and attaches holdings records to represent separate manifestations. In response to strong user and subject bibliographer preferences, New York University (NYU), routinely attaches all serial manifestations held by or accessible through the library to a single serial work or expression description. Each equivalent manifestation is then recorded and displayed through a separate MARC 21 holdings record. Every serial holdings record in NYU’s catalog contains at least the first two bytes of a 007 tag so the specific material designation (SMD) or carrier information of each is clearly displayed for users (e.g., text, online, CD-ROM, microfiche, and so on). Figure 4 represents an example serial record from NYU’s OPAC. Note the use of multiple “Library has” statements detailing numerous holding locations and formats directly on the print bibliographic description. In the MARC record, these “Library has” statements are generated through multiple 866 fields (Summary Holdings Statement). NYU’s ability to pursue this aggressive single-record technique for serial resources is facilitated by the Geac ADVANCE ILMS’s capability of attaching multiple serial receiving records to separate holdings records upon a single serial bibliographic description. In other words, for a serial title that NYU holds current subscriptions for print, online, and microfiche manifestations, ADVANCE enables our Serials Receiving Unit to order, receive, and check in the individual manifestation issues upon separate holdings records attached to a single bibliographic description. This receipt history is displayed in detail for OPAC users through the separate MARC 21 Holdings records attached to the single bibliographic description.

If librarians are prepared to reconsider the AACR community’s approach to cataloging manifestations and simultaneously demand revolutionary OPAC displays from ILMS vendors, it may be possible to avoid the requirement of exchanging complex, multi-tier records. When Attig and Yee independently proposed this idea several years ago, its realization seemed decades away. Yet computing has rapidly become so powerful, so ubiquitous, and so much less expensive that implementing such a system may be closer than we think. In his paper for Svenonius’s Conceptual Foundations of Descriptive Cataloging, Attig implicitly referred to the idea of a single, centralized catalog, envisioning an Elysian future wherein all catalogers would contribute to the same authority files and a single bibliographic catalog.

While most AACR2 libraries today take advantage of the centralized authority database represented by the LC/NAF, each library’s individual catalog, made up of contextually specific authority, bibliographic, and holdings records, remains completely disparate, and therefore isolated—connected and networked, but alone. As the AACR community moves toward implementing a cataloging code based largely upon the FRBR conceptual model and mindful of displaying the relationships among records and entities, this tension, as Attig calls it, between work input cooperatively and shared at the national or international level, and work that must then be replicated locally, will become increasingly redundant and frustrating.

In her contribution to the 1995 ALCTS preconference, “The Future of the Descriptive Cataloging Rules,” and again in her paper for the 1997 Toronto Conference, Yee sounds a more explicit call for a single shared catalog. The real problem with all linking devices in a shared cataloging environment, however, lies with the shared cataloging environment itself...
real solution . . . is that instead of sharing cataloging records, we need to reexamine the possibility of sharing a catalog! . . . If the development of the information superhighway eventually means cheap and ubiquitous telecommunication, could we not begin to envision a single catalog, accessible to all users, and updatable by all catalogers?87

The practical appeal and cost-effectiveness of a centralized, shared interface are difficult to ignore. During the serial rule revisions that followed the 1997 Toronto Conference and culminated with the revision of AACR2 Chapter 12 for Continuing Resources in 2002, the goal of just such a catalog was raised several times. Cooperative cataloging pushed to individual holding libraries is especially attractive for serials catalogers because of the ongoing bibliographic and holdings maintenance required by title changes and the issuance of serial resources over time. The centralized, distributed catalog envisioned by Attig and Yee would make such updating automatic. Each time users retrieved a serial title within the catalog envisioned by Attig and Yee would make such updating automatic. Each time users retrieved a serial title within this shared, centralized catalog, they would receive the most current bibliographic and holdings data available regardless of whether the latest updates were input locally or by another cataloger across the country or globe.

Further demonstrating how change remains the only true constant and just how quickly change occurs, recent merger announcements between OCLC and RLG and the subsequent consolidation of the Endeavor and Ex Libris ILMS systems certainly have far-reaching implications for library workflows and processing.87 Exactly what these combined interfaces may offer future librarians will take months or even years to determine. With a single, shared bibliographic utility in place though, the feasibility of this centralized, shared catalog interface remains one possibility.

Incorporating one of the MulVer solutions presented in this paper within this centralized catalog would produce an interface offering a win-win situation for all library players. Library administrators would like the lower cost structure, catalog librarians would feel empowered by entering real-time contributions in a single, shared catalog, and reference librarians and users would enjoy access to all available cataloged resources. Probably the only current players likely to be displeased with this new central catalog would be the ILMS vendors. Had ILMS vendors shown the initiative necessary to provide libraries with technologically enhanced ILMS systems and OPAC displays during the last fifteen years, libraries would not still be seeking solutions to display problems endemic to the automated catalog environment.

If many of these recommendations and proposals seem familiar, they should. In his 1989 paper titled, "Descriptive Cataloging Rules and Machine-Readable Record Structures: Some Directions for Parallel Development," Attig called upon the AACR and MARC communities to codify the necessary principles and to explore the systems design required to enable the cataloging code and the then newly developed USMARC Holdings format to resolve the MulVer problem.89

Schottlaender has discussed calls for Rule 0.24 reform on behalf of the AACR2 community dating back to the earliest multiple versions discussions.90 This reform movement reached a new high at the 1997 Toronto Conference where "it was clear that 'The Cardinal Principle' was a basic and pressing problem."91 As this paper illustrates, the JSC has now received similar messages from several user communities regarding this pressing problem for several years. The 1989 Multiple Versions Forum was a faint rumble. At the 1997 Toronto Conference, several papers and many presenters expressed continuing and mounting displeasure with AACR2's cardinal principle. Then, within fairly rapid order, two additional publications expressed dissent within the cataloging community: ISBD (ER) sanctioned multiple manifestations on single bibliographic descriptions in 1997, and the FRBR model in 1998 demonstrated an eagerness to consider overall catalogs in new ways with specific emphasis upon the needs of users.92 In 2003, the IFLA Cataloguing Section responded with a series of referenda in the form of International Meetings of Experts designed to solicit input and feedback on the feasibility of an internationally coordinated cataloging code.93 In something of a disappointment to librarians advocating the potential of FRBR and a more radical dismantling of the AACR2 Rule 0.24, the first International Meeting of Experts for an International Cataloguing Code (IME–ICC) held in Frankfurt among the European and American cataloging experts reaffirmed an insistent adherence to manifestation-level cataloging.94

As for the MulVer issue within the AACR community that today's primitive, manifestation-level OPAC displays perpetuate, this paper has explored three approaches to the problem, two long-term and another that could be explored and perhaps implemented more quickly. First, the revision of AACR and the eventual role Rule 0.24 may play within RDA is a long-term solution, for the expected publication date of the new cataloging code is 2009. Second, FRBR and its eventual impact upon the cataloging code are linked with this 2009 AACR/RDA timeline. Nonetheless, FRBR is already exerting influence on user interfaces and the future development initiatives ILMS vendors are considering. It seems quite likely that while the new cataloging code in 2009 will continue to instruct catalogers to build manifestation-level bibliographic descriptions, FRBR’s greater influence may be upon how ILMS system designers develop
OPACs to cluster these manifestation-level descriptions into work and expression-level displays for users.

The third and more immediate resolution to the MulVer problem resides with the MARC 21 communications formats. Libraries are constantly exploring this option within their local ILMS systems. Local resolutions to these ongoing problems are exemplified by initiatives at NYU and the UCLA Film and Television Archives. Within today’s cooperative cataloging environment containing shared bibliographic utilities, centralized authority files and distributed, separate institutional catalogs however, these initiatives stand out and, in some ways, prove problematic. These solutions are not perfect, but the current obstacles to fine-tuning them result more from a lack of development by ILMS designers in their indexing and OPAC display capabilities than from significant conceptual problems with work or expression-level displays.

Ultimate resolution to the MulVer problem resides with ILMS OPAC displays. For a number of practical reasons described in this paper, not least of which is the need to preserve the link between the OPACs of tomorrow and the millions of manifestation-level bibliographic records populating catalogs today, manifestation-level descriptions will remain the data packets libraries use to store and exchange records. The necessity for libraries to store and exchange data as cohesive manifestation-level descriptions though in no way forces OPACs to display data in the same way. ILMS vendors do so because it is easy and because librarians have not uniformly insisted they do otherwise. Librarians must cease this passive acceptance of the inferior OPAC displays bundled with today’s ILMS systems. Nonetheless, librarians must also bear partial responsibility for the failure of ILMS OPAC displays to develop further during the last twenty-five years. While it is easy to point the finger at library ILMS vendors, librarians have failed to present ILMS software designers with a cohesive vision of how OPAC displays should be improved. That time must end now. The pace of technological innovation across an array of professions and industries during the last fifteen years has been astounding. Libraries cannot afford to be left behind. Librarians must demand smarter displays from ILMS vendors, but they must be prepared to provide software designers with the direction necessary to develop such displays. Each of the millions of bibliographic and authority records in our catalogs represent rich data mines awaiting exploration and greater utilization. Yee’s recent analysis using existing MARC 21 records to generate work and expression identifiers in order to clarify OPAC displays for users is exemplary and should be required reading for ILMS designers, librarians, and library school students.55

Are these issues complex? Of course they are, but complex issues should not require UCLA’s Film and Television Archives to process moving image materials differently, or NYU to process serial resources differently than other library materials in order to fulfill their user’s needs. Historically, cataloging solutions within local settings have driven national and international policies. For example, the CONSER single-record approach and the aggregator-neutral record grew out of individual libraries solving complex problems for users in practical ways through local OPAC displays. It is time for librarians to determine if solutions to issues like the MulVer problem are complex because they have to be, or complex because librarians perpetuate practices that make them complex. What users need is simple. They need consistent access to content. Within today’s world of proliferating information carriers, providing consistent access to the content users seek is inherently complex, but to users it must appear simple. The job of today’s librarians is to apply complex solutions to attain apparent simplicity—call it the Zen of librarianship. For librarians to require or expect users to continue to learn or assimilate anachronistic procedures based on antiquated practices is unrealistic and threatens to render library catalogs and collections irrelevant. In fact, such expectations violate the purposes of the catalog formulated by Cutter and furthered by Lubetzky. In considering simple, consistent OPAC displays for users of our increasingly complex bibliographic catalogs, librarians and catalog designers would do well to consider the words of Dempsey:

The benefits of a more consistent OPAC display are clear: [Librarian’s] time and resources should be freed to think about collection and use of the collection, not consumed by the messy mechanics of acquisitions and processing; and the user experience should be shaped by learning and research needs not by the arbitrary constraints of interface and format. [Libraries] need to achieve the economies of consistent treatment as well as the benefits of consistent access.56

References


18. ALA Response to 5/JSC/AACR3/1 (draft of Part 1 of AACR3), 132. (As of Nov. 19, 2005, access to this document remained restricted.)


24. Ibid., 6.


28. Ibid.


32. Association for Library Collections & Technical Services, Cataloging & Classification Section, Committee on Cataloging: Description & Access Task Force on Rule 0.24, Overview and Recommendations Concerning Revision of Rule 0.24, 6.


34. Oliver, “FRBR is Everywhere.”


36. Howarth, “Content versus Carrier.”


39. Howarth, “Content versus Carrier.”


41. Ibid.

42. Wilson, “The Second Objective.”


44. John C. Attig, e-mail message to the author, July 9, 2005.

47. Ibid.
48. Ibid.
51. Antelman, “Identifying the Serial Work.”
52. Ibid., 249.
61. Ibid., 183.
64. Coyle, “Future Considerations”; Mimno, Crane, and Jones, “Hierarchical Catalog Records.”
76. International Federation of Library Associations and Institutions UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), Functional Requirements for Authority Records [draft 2005/06/15].
81. Ibid.
82. Ibid., 6.
84. Attig, “Descriptive Cataloging Rules and Machine-Readable Record Structures: Some Directions for Parallel Development.”
85. Ibid.
89. Attig, “Descriptive Cataloging Rules and Machine-Readable Record Structures.”
91. Ibid., 17.
95. Yee, “FRBRization.”
2007 LITA National Forum

Come join us in the “mile-high” city of Denver to celebrate 10 years of the LITA National Forum.

Don’t miss these great Keynote Speakers at the LITA National Forum:

- **Jeffrey Kiehl**  
  National Center for Atmospheric Research, CO

- **David King**  
  Topeka & Shawnee County Public Library, KS

- **Jeremy Frumkin**  
  Gray Chair for Innovative Library Services, OR

Plan now to join us in downtown Denver for the Forum: a three-day educational event that includes preconferences, general sessions, and more than 30 concurrent sessions.

Register Now!

www.lita.org/forum07

October 4–7, 2007
Cross-lingual Name and Subject Access: Mechanisms and Challenges

By Jung-ran Park

This paper considers issues surrounding name and subject access across languages and cultures, particularly mechanisms and knowledge organization tools (e.g., cataloging, metadata) for cross-lingual information access. The author examines current mechanisms for cross-lingual name and subject access and identifies major factors that hinder cross-lingual information access. The author provides examples from the Korean language that demonstrate the problems with cross-language name and subject access.

Today’s global information society, benefiting from rapidly advancing communication technologies, spans geographical, lingual, and cultural boundaries. Recognition of the need for knowledge organization and integration, and access to cross-cultural and cross-lingual resources has greatly increased. The 2004 ISKO International Conference on “Knowledge Organization and the Global Information Society” and a 2004 special issue of Cataloging and Classification Quarterly (“Knowledge Organization and Classification in International Information Retrieval”) are two examples.1 International digitization projects have opened access to medieval texts as well as images and primary sources housed in libraries and institutions around the world, greatly advancing global access to multicultural resources.

The technological revolution that brought forth the global information society also has spurred recognition of the necessity for international collaboration aimed at multicultural education and diversity.2 Linguistic and computational linguistic communities have collaborated in developing multilingual information resource discovery tools, such as concept-based indexing. These are used primarily for cross-lingual information processing. One example is EuroWordNet, which is based on Princeton University’s WordNet, a lexical database for the English language.3 The Open Language Archives Community (OLAC) has also been engaged in archiving, disseminating, and preserving language and cultural resources, including language-engineering tools, through utilization of the Dublin Core metadata standard.4

The challenges of accessing resources across cultures and languages suggest this is an area of particular interest to librarians, who are responsible for description and access. As a first step in exploring this topic, the author studied current practices in providing cross-cultural and cross-lingual information access. In this paper, she identifies problem areas and suggests directions for future study. The scope is limited to studies dealing with cataloging and metadata schemes for cross-cultural and cross-lingual information access.
Approaches to Cross-lingual Information Access

The development of cross-lingual thesauri, subject heading lists, and name authorities, as well as the translation of the Dublin Core (DC) metadata scheme into many different languages, is ongoing. In addition to the activities of the DC Metadata Initiative for developing multilingual DC metadata, various approaches to building cross-lingual knowledge organization schemes have been developed with an eye to better access to multicultural and multilingual resources.\(^5\)

Language engineering and linguistics communities have developed lexical tools for cross-lingual resource discovery; these include machine translation, ontology, information extraction, text summarization, and speech processing. Multilingual information resource discovery tools such as concept-based ontology (e.g., EuroWordNet and Global WordNet Association) also have been developed.\(^6\) OLAC has been engaged in archiving, disseminating, and preserving language-culture related resources by developing the OLAC Metadata standard, which defines the format used for the interchange of metadata within the framework of the Open Archives Initiative (OAI).\(^7\) The metadata set is based on the complete set of DC metadata terms, but the format allows for the use of extensions to express community-specific qualifiers.

In library communities, cataloging and metadata standards have been internationalized. Cross-lingual subject access via conceptual mapping of Library of Congress Subject Headings (LCSH) and cross-lingual name access through cross-linking of Library of Congress (LC) name authorities have been undertaken. The following sections present a literature review and identify the challenges inherent in transliteration and word segmentation in nonroman scripts, with particular attention to Korean. Challenges in building subject heading and name authority files for cross-lingual information access also are discussed.

Cross-lingual Subject Access: Conceptual Mapping Mechanisms

Heiner-Freiling reported the results of a survey of national libraries on subject headings conducted under the auspices of the International Federation of Library Associations and Institutions (IFLA).\(^8\) According to the survey data, LCSH is predominantly used in twenty-four national libraries of English-speaking countries; in addition, a translated or modified version of LCSH is being used in twelve other countries. Several authors have written on the problems caused by translated subject headings across languages and cultures.\(^5\)

Subject headings of Korean collections in North American libraries are based largely on LCSH, a translation from the source language (i.e., Korean) into LCSH in English. This author presented an earlier analysis of the problems in subject headings translated between English and Korean.\(^9\) Problems that occur in translated subject headings likewise can be expected to occur in any metadata mapping process between the two languages.\(^10\)

The concepts of LCSH are formulated into various syntactic forms—single noun, compound noun, noun phrase, and inverted phrase. The concept of a heading can be expressed in several different forms, leading to potential complexities and inconsistencies. Partially due to the multiple morpho-syntactic forms used in expressing the same concept, cataloger inconsistencies exist even when working with a single language, such as the assignment of subject headings in English by an English-speaking cataloger to works in the English language. The translation process between two languages only exacerbates such inconsistencies.

Korean subject cataloging suffers from the inevitable drawbacks of assigning Korean concepts by employing English subject headings. The conceptual mismatch and difficulties of translation from one language to another are largely due to different linguistic structures and socio-cultural norms. In the case of English and Korean, these structural differences are considerable, unlike between English and Spanish, because English and Korean are unrelated languages. For example, Korean is an agglutinative language in which functional particles, such as case markers and functional affixes, are attached onto the content words as grammatical operators. On the other hand, English and Spanish lack such characteristics. Instead, they are heavily dependent on word order to designate grammatical function. The manner of conveying a semantic concept may be manifested differently in Korean and English language users. Such differences in conceptual manifestation are greatly increased in the process of translation.

The following example of a translated subject heading exemplifies these problems. The romanized Korean compound phrase *Hanguk mal* could be translated as:

\[
\begin{align*}
\text{A Korea language/} & \text{The Korean language} \\
\text{The language of Korea/l} & \text{anguage of Korea} \\
\text{Korea and a language/} & \text{Korea and languages}
\end{align*}
\]

The Korean heading may be translated into English with various forms. Major differences among these possible headings include the following: the prepositional phrase *The language of Korea* and the conjunctural phrase *Korea and languages* show indefinite and definite article variants (a versus the) and inflectional variants (language versus languages). Written Korean employs grammatical devices...
such as particles (e.g., case markers denoting subject and object) and suffixes. These can be omitted in the spoken form without causing any communicational ambiguities. In the written language, as with *Hanguk mal* in the previous example, the omission of such functional words readily gives rise to ambiguity: *Hanguk ui mal* is translated as the prepositional phrase The language of Korea. On the other hand, *Hanguk kwag mal* is translated as Korea and languages. Thus, omission of the grammatical particles *ui* “of” and *kwa* “and” creates conceptual ambiguity.

Kwasnik and Rubin examined challenges in conceptual translation of classification schemes across languages and cultures. They assessed differences in kinship terms in fourteen languages, revealing the challenges and problems inherent in the process of translation of a classification system. As a framework for culturally sensitive classification translation, certain modifications to the classification system (adding or deleting terms or both) reflect individual linguistic and cultural characteristics and are inevitable. In the case of one-to-two mapping, creation of cross-references is a practical step forward in clarification. In a similar manner, the use of modifiers or scope notes in order to avoid conceptual ambiguity would be advisable.

### Multilingual Access to Subjects: Cross-linking Mechanisms

To date, the major project on multilingual subject headings has been Multilingual Access to Subjects (MACS), which aims at providing English, French, and German subject access in library catalogs through cross-linking techniques. Clavel-Merrin, MacEwan, and Landry reported on this project. The project has been conducted by European national libraries—the Swiss National Library, the Bibliothèque nationale de France, The British Library, and Die Deutsche Bibliothek—through international collaboration under the auspices of the Conference of European National Librarians. The cross-linking technique is based on conceptual mapping among the authorized headings of three subject lists: English—LCSH, French—RAMEAU (Répertoire d’autorité matière encyclopédique et alphabétique unifié) and German—SWD/RSWK (Schlagwortnormdatei/Regeln für den Schlagwortkatalog). Through a manual cross-linking process, conceptually equivalent linking is established. If no equivalent concept exists across the three subject headings, the heading stands alone.

The project began with a subset of headings in the areas of theater and sports. The rationale for selecting those areas was to test universality and cultural variation. The area of sports would be expected to have a high conceptual correspondence across the three languages and the three subject heading lists because the area of sports is considered to be a less culture-bound domain; conversely, the area of theater reflects culture-specific terms and concepts and low correspondence across these subject headings would be expected.

As expected, cross-linking in the area of sports yielded a high degree of equivalence. MacEwan reported that when comparing terms in a sample of 278 sports subject headings, 86 percent of headings matched across all three subject headings lists, 8 percent of headings matched across two lists, and 6 percent of headings were unmatched. In the more culture-bound domain of theater, the cross-linking match was much lower than in the less culture-bound domain of sports. MacEwan reported that, when comparing terms in a sample of 261 theater subject headings, 60 percent of headings matched across all three subject heading lists, 18 percent matched across two lists, and 22 percent of headings were unmatched. A concept realized as a word in one language can be equivalent to a linguistic morpheme (the smallest unit of meaning in oral and written language), word, phrase, or clause in other languages. Thus, syntactic variations are expected to hinder the mapping process. MacEwan gave an example of the challenge seen in creating a conceptual linking system across three subject headings (English, French, and German) in the following: “Track athletics—Coaches in LCSH matches with Leichtathletiktrainer in the SWD, but in RAMEAU it is only matched by adding a subdivision to the authority record at the point of indexing a document: Athletisme-Entraineurs.” To alleviate mapping problems caused by such syntactic variations, links between headings and strings are allowed. In addition, the creation of new headings is allowed to create a conceptual mapping between the subject heading lists, as long as there is literary warrant in the catalog of the user institution.

### Conceptual Mismatch between Target and Source Languages

The conceptual mapping process is analogous to translating two or more different languages. Figure 1 illustrates some possible conceptual mismatches in the process of semantic mapping between two languages. Precise and equivalent mapping between two languages in translation does not exist. The first and second diagrams in figure 1 illustrate the necessity for strategies to deal with inexact equivalence in the case of one-to-many and many-to-one mapping. In the case of no conceptual equivalence, shown in the third diagram, the general concept in the target language might serve as an alternative for semantic mapping. However, due to the lack of specificity, the alternative general concept may not contain the original source concept, resulting in an unavoidable limitation in cross-linguistic situations.
Owing to the dramatically different language structures and cultural bases of Korean and English, translated subject headings involving these languages frequently are not equivalent to the concept of the original heading. The concept of the translated headings is either overly broad or the headings do not retain the original meaning. Thus, a more thorough analysis and understanding of the very different Korean and English language structures are needed to alleviate this inevitable difficulty.

The subject heading that follows, taken from a MARC record describing a Korean monograph, *pumasi*, illustrates the challenges faced in conveying the original concept in the process of mapping from the Korean concept of a word to LCSH.

```
650 0 Interpersonal relations.
651 0 Kyonggi-do (Korea)$x social life and customs.
```

The title of the book is *pumasi* (exchange of services/labor) *wa* (and) *chong* (affection) *ui* (of) *ingan* (human) *kwangye* (relationship). The translation could be *The interpersonal relations of the exchange of labor and affection.*

The word *pumasi* describes the social structure of Korea in the agricultural context. The *pumasi* is the system by which people effectively provide help to one another. People who are in need can obtain financial and other help from others for a short period without paying interest. They will return the *pumasi* on some other occasion when the people who gave help are themselves in need of help. This system was originally developed in a traditional agricultural society and then transferred into the urban society of modern Korea. The underlying concept of *pumasi* may be stated thus: *solidarity with affection in a community.*

LCSH does not have a heading that is equivalent to the *pumasi* system. This is because *pumasi* is a product of Korean culture. In order to denote the subject heading, then, a broad and general heading such as *social life and customs* would be employed for this monograph in the topical subdivision of the heading (i.e., 651). As can be seen, the translated subject heading in the above record loses the original concept of the Korean heading due to conceptual mismatch.

### Cross-lingual Name Access through Cross-linking Mechanisms

Two major projects on cross-lingual name access through the cross-linking mechanism utilizing roman script currently are employed. One is the Virtual International Authority File (VIAF), a joint project between LC and Die Deutsche Bibliothek, with OCLC’s research support. VIAF is a single personal name authority file that combines the name authority files of both institutions through the cross-linking mechanism.

In the VIAF project, the authority records from Die Deutsche Bibliothek are matched to the corresponding LC authority records through the cross-linking mechanism. Following this linking process, maintaining the authority files and providing user access to the files will be through the shared OAI servers. Upon the completion of the project, each user group in the United States or Germany will be able to view personal name records established by the other institution and view the personal name records of each user group’s own language.

The other project dealing with roman script is Linking and Exploring Authority Files (LEAF), which was established in 2001 with the involvement of fifteen organizations utilizing eight languages. Clavel reported two principal challenges in establishing a cross-lingual authority file. Both challenges are derived from linguistic variation and ambiguities across languages. First are language-specific features such as the order of components in compound names, location of particles, and numbering system for kings and popes. The second challenge concerns standardization of methods for disambiguation of homonyms. Natural lan-
language is full of lexical ambiguities. For instance, homonymy creates ambiguity (e.g., bank [building] versus bank [river]). Homonyms have the same lexical form but manifest unrelated meanings that are arbitrarily developed. The Anglo-American Cataloguing Rules, 2nd edition (AACR2) chapters 22 through 26 present pragmatically constrained disambiguation techniques for the names of persons, corporate bodies, and places by differentiating contexts. For example, to disambiguate identical names, birth or death dates (or both) are added (e.g., John Q. Smith [1904–1972] versus John Q. Smith [1905–]). In the case of ambiguous corporate body names, a qualifier is added—e.g., John Smith (firm). According to Clavel, the addition of academic and nobility titles is generally standardized for disambiguating homonyms. The specification of profession or activity, however, is much less standardized. Accordingly, this creates problems in cross-linking of authority files across languages.

Several authors have looked at nonroman scripts (particularly East Asian languages such as Korean, Japanese, and Chinese) and have found that transliteration causes cross-lingual name access problems, because of the nature of the language. Names in the Korean, Chinese, and Japanese languages utilize Chinese ideographs owing to a common history; thus, variant forms of names are represented in these languages. For example, in the case of the Korean name, Hangul (Korean vernacular script), Chinese ideograph and the transliterated form are all used.

When discussing Korean, one must take into account the differences in transliteration schemes between those based on phonetic structure and those based on morphemic structure. Differences in transliteration schemes are also applicable to other nonroman scripts.

For instance, LC’s relatively recent adoption of the Pinyin transliteration scheme from the Wade-Giles scheme in transcribing Chinese language materials illustrates the complex issues surrounding the differences in transliteration schemes even involving the same language. Arsenault reported on an experiment in retrieval efficiency among monosyllabic Pinyin, polysyllabic Pinyin, and Wade-Giles while searching known item exact title and keywords in title. The findings of the study demonstrate that the poly-syllabic Pinyin system, which transcribes Chinese according to syntactic unit (i.e., word by word), significantly increases retrieval efficiency compared to monosyllabic Pinyin and Wade-Giles, which share the feature of transcribing Chinese morpheme by morpheme.

Naito presented a variety of ways of transcribing the same Japanese name, such as phonetic transcription in Hiragana and phonetic transcription in Katakana, transcription in simple form, and Chinese scripts. Table 1 (from Naito) illustrates this.

This author presented issues relating to the Korean transliteration scheme. In South Korea, no unified transliteration scheme is used. Different transliteration schemes are employed in different sectors for varying uses. For example, libraries and publishing industries employ the McCune-Reischauer (MR) system in publication and bibliographic records. The Yale system is uniformly used by linguists within Korea and abroad. Lastly, government documents, including street signs and road maps, employ the Ministry of Education system.

The differences among these schemes reflect the linguistic representation of sound systems. The MR system and Ministry of Education system are based on the phonetic structure of Korean. Transliteration based on phonetic structure encodes words in the manner in which they are pronounced. For example, in English the word two is transcribed phonetically as [tu].

The Yale system is based on morphemic structure. Morphemic structure-based transliteration transcribes the base form of a word regardless of sound changes. Korean is a language that employs rich morpho-phonemic complexity. The base form of a word changes according to the adjacent sound environment. Most agglutinative languages, including Japanese, fall into this category. They are all very complicated morpho-phonemically. For example, the form of the Korean word mul (water) is changed into muri when the subject case particle -i is attached to it. Morphemic structure-based transliteration is not reflective of sound change as is the phonetic type of transliteration utilized in the MR scheme; instead, it reflects the base form.

The current cataloging system dealing with Korean materials employs the MR transliteration scheme. One of the major drawbacks of the use of the MR system is that it causes semantic loss. This is especially critical in the area of name access. Transliteration of words following the way in which they are pronounced has the potential of representing

<table>
<thead>
<tr>
<th>Table 1. Japanese personal name</th>
</tr>
</thead>
<tbody>
<tr>
<td>黒澤明</td>
</tr>
<tr>
<td>赤澤明</td>
</tr>
<tr>
<td>くろさわあきら</td>
</tr>
<tr>
<td>クロサワアキラ</td>
</tr>
<tr>
<td>クロシフアキラ</td>
</tr>
<tr>
<td>Kurosawa Akira</td>
</tr>
<tr>
<td>KUROSAWA Akira</td>
</tr>
<tr>
<td>Akira Kurosawa</td>
</tr>
<tr>
<td>Akira KUROSAWA</td>
</tr>
</tbody>
</table>

a name ambiguously. For example, the Korean name Kim Sok-min becomes Kim Song-min according to the MR system. With author names transliterated according to the MR system, ambiguity becomes almost inevitable. The linguist Ramsey noted that “This information loss becomes especially critical when all cataloging work is done by computer, and so it is perhaps time to give some thought as to how appropriate McCune-Reischauer is in cases where precise data processing is required.”

The MR system, based as it is on phonetic structure, does not disambiguate different meanings of homographs (i.e., same words but different meanings), one of the primary causes of semantic ambiguity. This phenomenon can be illustrated by an example in English: two, to, too. If these three lexical items are transcribed according to the pronunciation [tu], the resulting semantic ambiguity can be clearly seen. This happens frequently with the MR scheme. Such ambiguities inevitably cause significant impediments in the process of information retrieval.

In addition, the MR system results in variations in the creation of bibliographic records. When catalogers transcribe words according to pronunciation, they can create inconsistent and arbitrary records. This is based on the fact that the pronunciation of words can vary according to speech style. If a cataloger pronounces a word or phrase using careful speech style, the resulting transcription would be different from that of a transcription based on casual speech style. The creation of differing bibliographic records is thus entirely possible, either by the same cataloger or different catalogers transcribing identical material.

The following bibliographic record illustrates this problem.

100 1 Kim, Young-un,$c 1927-
245 10 Ceh-2 $k^onggungnon : $bkungmin kukka $ui wans$ong $ul wihay$o
   /$cKim Yong-un.
246 3 Ch”io”an.
260 S^oul T”^ukpy^olsi :$Chisik San^opsa,
   $c1998.

The portion of the title field (245) in bold, k^onggungnon, reflects the casual speech style. If the cataloger who created this record had pronounced it using careful speech, the final consonant of the first syllable (i.e., kon) remains as a nasal sound, as indicated in bold: k^onggungnon, as opposed to k^onggungnon. In casual speech, however, the nasal sound [n] becomes assimilated into the following velar sound [ng].

The MR transliteration scheme contains inherent inconsistencies that can have a significant impact on information organization and retrieval. Semantic ambiguity, inconsistency, and semantic loss are critical issues hindering information retrieval and sharing bibliographic records. Consequently, the goals of bibliographic control are not achieved.

Problems of Word Segmentation

Difficulty in word segmentation occurs in agglutinative languages such as Japanese and Korean because of their inherent morpho-syntactic flexibility. Agglutinative languages allow functional particles such as case markers and inflectional affixes to be attached onto the content words as grammatical operators. For example, the word muli [water + subjective case affix] is composed of the content word (i.e., mul: water) and the functional affix (i.e., i: subjective case marker). This creates flexible word segmentation between functional and content words. Such flexibility of word segmentation in Korean creates inconsistent and arbitrary practice in word division; such inconsistency can be found in even the most authoritative Korean dictionaries. According to Yi Sung-u, word segmentation errors appear in 20 percent of Korean standard books in the school system.31 This highlights the difficulty in conducting word segmentation in the written Korean form.

Arbitrary word segmentation does not cause communicative problems in everyday language use, since communicative ambiguities stemming from inconsistent word segmentation can be resolved through contextual cues. However, such flexibility in word segmentation is a critical factor in hindering information sharing and discovery in the digital environment, which does not provide contextual cues.

The Library of Congress ALA-LC Romanization Tables provides rules specifying word segmentation and offers four basic underlying principles.32 The first basic principle is “Each word or lexical unit (including particles) is to be separated from other words.”33 The following Korean bibliographic record illustrates this principle.

245 00 Y^oksa sok $ui in’gan kwa chis^ong
   $ul t”^angu handa $/c Kim Chae-yong
   . . . [ et al. ] $p’y^on.
250 Che l-p”an.
260 S^oul $/bHan’gilsa,$c 1998.

The title field (245) can be segmented in the following way: Yoksa sok $ui in’gan kw^a chisong $ul t’angu $handa. The segmentation is denoted by the mark $, designating a total of eight word divisions. This principle follows one of the suggestions presented at the 1981 workshop conference on Korean transliteration, held at the University of Hawaii under the auspices of the Korean Studies Center, and reported by Austerlitz.34 The main aim of the
conference was to examine the Korean transliteration system (i.e., MR system) to produce consistent guidelines for transliterating Korean language.

This principle creates problems when users search a bibliographic record because word division following the LC principle is not utilized by Korean users; it is contrary to conventional practices of the language. The previous example title consists of only three word divisions in the Korean written form: Yokasokut^ in^gankwa^ chisongul^ t^anguhanda. Moreover, this rule presents another intrinsic difficulty. It applies only to case particles of a noun phrase, not to affixes of verb phrases. Thus, the word division principle is not applied to entire units of the sentence.

The MR transliteration scheme based on phonetic structure has critical drawbacks because it causes semantic loss, semantic ambiguity, and cataloging inconsistency. A transliteration scheme based on morphemic principles has substantial merit because it significantly contributes to resolving semantic ambiguity and inconsistency. One of the principal advantages of basing transliteration on morphemic principles is that the need for diacritical symbols also is substantially reduced, in contrast to a transliteration scheme based on phonetic principles, which increases the employment of diacritical symbols.

Word segmentation in agglutinative languages is very flexible. Even though guidelines and rules for word division exist, inconsistent and arbitrary practices are inevitable. An automatic parser of word segmentation based on linguistic principles is critically needed to ensure consistency of bibliographic records.

Linguistic Universality and Relativity across Language Structures

Impediments to enhancing access to cross-cultural and cross-lingual resources are largely derived from the complexities and variation of linguistic structures across languages. Linguistic and cultural approaches in developing cross-lingual and cross-cultural knowledge organization systems are critically needed.

The facility of natural language, in all its complexity, variability, and richness, is the defining aspect of humanity. This very complexity of expression and richness of lexicalization and linguistic structures becomes problematic in the electronic environment of information retrieval. Even though natural language possesses some characteristics that are independent of a specific language, many more language-specific characteristics exist. Such language-specific characteristics demonstrate that the structure of language is so closely intertwined with its source culture and society that it is inseparable from it. Natural language is not just mere arrangements of words, but the mirror of culture.

Combinations and arrangements of words do not reflect specific cultural and pragmatic meanings that are inherent characteristics in any given language structure.

Language-specific variations and differences in lexicalization patterns can be found easily in everyday language uses such as naming conventions, kinship terms, address forms, numbering systems, color terms, and names for body parts. For example, in Anglo-American society, building designs (e.g., LeBow College of Business), brand names (e.g., Ford), and even common reference nouns (e.g., maverick, boycott, lynching) originating from family names or titles are common. Conversely, this phenomenon is nonexistent in Korean language and society. Thus, one can say that this English-specific naming convention manifests the cultural trait of Anglo-American society.

Collectivist-oriented cultural and social norms, based on hierarchical structure, are closely reflected in the Korean language. This can be especially seen in the sophisticated honorific system and in the employment of various linguistic devices, such as lexical items existing in both plain and honorific form (e.g., na/chro [plain/honorific form] T, na/yonsey [plain/honorific form] ‘age’, chada/chumusida [plain/honorific form] ‘sleep: verb), to name a few. It is also seen in syntactic structures (e.g., honorific agreement in subject/object, predicate, and case markers). Such variant lexical forms are merely one illustration of a synonymy phenomenon that is not found in English, as shown in table 2.

The Need to Develop Interoperable Guidelines for Cross-linking Names and Subjects and Conceptual Mapping

A critical need for the development of common guidelines for cross-linking of names (e.g., person, place, corporate body) across languages exists. Development of such interoperable cross-linking guidelines should be guided by the examination of morpho-syntactic variations across language structures, especially for the structures of names.

Word segmentation and transliteration schemes dealing with nonroman scripts also play a part in limiting access to cross-lingual and cross-cultural resources. Standardization of such transliteration schemes and development of mechanisms geared toward consistent word segmentation also are critically needed. Specifically, reexamination of transliteration schemes and development and application of a morpho-syntactic parser based on linguistic principles for automatic word segmentation are vital conditions for cross-lingual information access.

Development of knowledge organization schemes for cross-lingual subject access also is hindered by the lack of common conceptual mapping criteria that are interoperable across languages and cultures. Semantic mapping, involv-
Cross-lingual Name and Subject Access

ing metadata and subject heading lists across languages, is one of the most critical issues in resource discovery and information exchange. Without achieving interoperability of semantic mapping, application of cross-lingual knowledge organization tools for the retrieval of networked resources will be significantly hindered. In order to develop interoperable conceptual mapping guidelines across languages and cultures, identification of lexicalization patterns based on semantic, syntactic, and pragmatic linguistic analysis is critically needed.

Cross-linguistic differences result in conceptual and lexical gaps and overlaps between target and source languages that present themselves during the mapping process. Conceptual mapping between languages presents a variety of lexical gaps and overlaps including inexact equivalence, partial equivalence, nonequivalence, and single-to-multiple equivalence. Culture-specific language characteristics suggest that, in order to overcome problems in the development of cross-lingual knowledge organization tools (e.g., subject headings, thesauri, metadata) and to ensure interoperability among these tools cross-linguistically, language-specific characteristics must be taken into account.

Table 2. Korean lexical honorific system

<table>
<thead>
<tr>
<th>English equivalent</th>
<th>Plain form</th>
<th>Honorific form</th>
</tr>
</thead>
<tbody>
<tr>
<td>age: noun</td>
<td>nai</td>
<td>yonse</td>
</tr>
<tr>
<td>house: noun</td>
<td>chip</td>
<td>taek</td>
</tr>
<tr>
<td>sleep: verb</td>
<td>chada</td>
<td>chumusida</td>
</tr>
<tr>
<td>eat: verb</td>
<td>mokta</td>
<td>tusida</td>
</tr>
</tbody>
</table>


Complexities and variations of linguistic structures across languages and cultures have a significant effect on name and subject access across languages. Thus, study of linguistic and cultural approaches to developing cross-lingual and cross-cultural knowledge organization systems is critically needed. The major research gaps in current literature concern addressing issues in relation to developing interoperable guidelines for cross-linking of names and developing common conceptual mapping criteria that are interoperable across languages and cultures for cross-lingual subject access.

Conclusion

Are multiple subscription sources multiplying your subscription problems?

We can help.

EBSCO brings simplicity and efficiency to the management of subscriptions, from order placement to renewal. As your single-source supplier for print, e-journal, e-book and database subscriptions, EBSCO handles all ordering and countless access details on your behalf.

Our proven approach saves your staff time and your library money by eliminating the need to communicate with multiple contacts and deal with assorted reporting methods. EBSCO’s subscription management service provides you with realistic, helpful solutions.
This underlies the necessity of future studies in morpho-syntactic variation across languages for cross-lingual name access and an examination of lexicalization patterns based on semantic, syntactic, and pragmatic linguistic analysis for cross-lingual subject access. Drawbacks in word segmentation and transliteration schemes dealing with nonroman languages also call for reexamination of transliteration schemes and for the development of a morpho-syntactic parser for automatic word segmentation.

References


16. Ibid.
17. Ibid., 203.
22. Clavel, “LEAF.”
33. Ibid., 82.
This study examines problems caused by initial articles in library catalogs. The problematic records observed are those whose titles begin with a word erroneously considered to be an article at the retrieval stage. Many retrieval algorithms edit queries by removing initial words corresponding to articles found in an exclusion list even whether the initial word is an article or not. Consequently, a certain number of documents remain more difficult to find. The study also examines user behavior during known-item retrieval using the title index in library catalogs, concentrating on the problems caused by the presence of an initial article or of a word homograph to an article. Measures of success and effectiveness are taken to determine if retrieval is affected in such cases.

When filing entries alphabetically in an index, ignoring initial definite and indefinite articles is customary. For instance, the book titled The Earth and Its Inhabitants is normally filed under the letter “e.” This procedure is used almost universally because initial articles “tend to be used intermittently,” and also because, due to the high occurrences of initial articles in titles, it would otherwise produce very large groupings of entries beginning with the same word, thus losing the desired alphabetical dispersion of entries within the index. In the current version of the MARC 21 standard, this procedure can be achieved, for the first index subfield in some fields, by using a numerical indicator (the non-filing characters indicator) corresponding to the number of initial characters to be ignored for indexing. Using this technique allows the initial article to be retained in the title field and used for display, without being taken into account in the browse index.

Because the non-filing indicator is not available for all the fields in which articles and other non-filing elements occur, and also because non-filing data elements do not always occur at the beginning of a field, a new technique, setting off the non-filing zone by means of control characters, was approved in 1999 as a result of American Library Association (ALA) Machine-Readable Bibliographic Information (MARBI) Committees Proposal 98-16R. Guidelines for use of the new non-filing control characters were discussed in two discussion papers, DP118 (June 1999) and 2002-DP05 (January 2002), and finally published in...
2004 by the Network Development and MARC Standards Office of the Library of Congress. This procedure offers more flexibility, as it allows the cataloger to identify non-sorting zones virtually anywhere in the record and tag them with the use of special control characters whose function is to delimit the beginning and the end of the non-filing elements. As far as data representation is concerned, there are fairly standardized, documented, and efficient ways of dealing with initial definite and indefinite articles in data elements; however, the MARC coding controls only the way initial articles are to be indexed, not the way the retrieval is done. Less standardization is found at the retrieval stage and this is what is investigated in this study.

All systems preprocess search strings to some extent (e.g., ignoring case distinction, omitting punctuation or replacing it with spaces, ignoring diacritics) before sending them to the index. When a user launches a browse-title search in a library catalog, the retrieval module may activate an algorithm to detect the presence of an inopportune initial article at the beginning of the query string. Because most initial articles are removed from the entries when indexing the title strings, even if a user includes an initial article in his or her query, the algorithm will automatically eliminate the word/article and bring the user to the correct entry point in the index. This procedure may prove very useful in some cases. For instance, if the user retains the initial article in a search query (for example, ti=the earth and its inhabitants), the algorithm detects the initial article and automatically suppresses it from the search query before it is sent to the index. In this example, the system therefore will bring the user the index of titles beginning with the letter “E” rather than the letter “T.”

Nonetheless, most of these algorithms are not sophisticated enough to detect some linguistic subtleties, which can result in retrieval problems. This automatic detection of initial articles in search queries poses a number of problems, particularly in multilingual environments. The cataloger’s decision to declare an initial word as an article to be ignored must be based on several factors, among which the language comes first, since it can be reasonably assumed that an initial article in one language will have a corresponding legitimate non-article equivalent in another language. This is the case, for instance, in German with the article “die,” which is homographic to (i.e., spelled with the same sequence of letters as) the English verb “to die.” It would not be correct to file the title Die Another Day under the letter “A.” In some cases, it is even necessary to grammatically analyze the titles in order to avoid incorrect assumptions within a language. In French, for instance, the definite article “la” is homographic (albeit the diacritic) to the adverb of place “là” (there); and the word “un” can either be an indefinite article, as in Un destin tragique, a pronoun, as in L’un d’entre eux, or a number, as in Un, deux, trois, partez! It can even be part of an adverbial locution, as in Un peu de fatigue. That is not counting the fact that it also is the homograph of the acronym form for United Nations (UN). Therefore, processing titles case by case is essential. Also, sentences (and titles) can begin with only one article, so it makes no sense grammatically to remove two or more words from the beginning of a title search query. Yet, the algorithms tested in this project will remove any number of words that appear at the beginning of a search query that match the words in their exclusion list. For instance, in Atrium (the Université de Montréal catalog), the query “un thé au Sahara” will be transposed to “au sahara” because the “un” matches a French article and the “thé,” when transposed to “the,” matches an English article.

The detection algorithms included in most information retrieval systems are not sophisticated enough to detect these linguistic subtleties, which are the cause of some retrieval problems. Some homographic non-article words might be erroneously removed from the queries. This is the case for a title such as Las Vegas, The Success of Excess. This title will be correctly filed in the index under letter “L,” since the word “Las” is part of a place name, but if the word “Las” is included in the exclusion list of the algorithm, it will be interpreted as the Spanish definite article and automatically stripped of the query string, and the user will be misguided to the letter “V” in the index where the entry is nowhere to be found.

Suppose a user needs to find the work by Michel Leiris entitled À cor et à cri. Browsing through the title index normally would be done with the standard query “a cor et a cri.” Unfortunately, if the initial article detection algorithm is activated, the user will be misguided to letter “C” in the index since the initial “a” of the query will be, in this case, wrongly interpreted as the English indefinite article “a” and the query text will be truncated, often without the user being aware of it, becoming “cor et a cri.” The title having been correctly indexed under letter “A,” the user will be wrongly positioned in the index as illustrated (figure 1) and may wrongly assume that the title is not in the collection. This lack of system feedback most probably has a negative impact on end users learning to use the catalog.

In the catalog (the University of Toronto catalog) in figure 1, the user has to choose between two search modes: either the keywords mode (containing), or the browse mode (starting with). If the starting with option is chosen, the user will probably draw the conclusion that the document being sought is not in the catalog, since the title is not displayed in the results. The record nonetheless can still be retrieved using a keyword search. Choosing the containing option presents another difficulty. In the catalog of the University of Toronto (in April 2006), querying “cor” produces 1,160 records, which must then be painstakingly examined one by one; querying “cri” produces 219 records, which is better.
This is still high, especially considering that the search is for a single known title. Querying “cor cri” (with an implicit Boolean AND) produces five records, which is more acceptable. Nonetheless, some titles only offer very limited terms when searched in the keywords mode—for example, À la française or À tous. Such searches in keywords mode lead to very large search results sets that are virtually unusable—6,608 and 2,093 results respectively (in the University of Toronto catalog).

A more efficient solution may be to deactivate the initial article detection algorithm in the search module and to replace it by providing the end users with clear instructions on omitting initial articles in queries. Taylor reports that if the instructions are clearly positioned (see figure 2 for an example) users will follow the instruction: “users tend to follow this advice if the instruction is noticeable and can be seen from the search box.”

Given these observations, one may question the usefulness of an initial article detection algorithm based on an exclusion list in a library catalog since its use may cause as many problems as it solves. On the one hand, the use of an exclusion list affords some help to the naive searcher by participating in the formulation of his or her queries. Such users thus can cut the electronic reference retrieved and paste it directly in the Search dialog box of the catalog without concerning themselves about anything else. If the title begins with a word or a series of words that are contained in the exclusions list, the search algorithm will remove the unnecessary words from the query without a user’s knowledge. On the other hand, this very exclusion list has several drawbacks and can disadvantage the users. One may ask, therefore, what course to follow. A professional librarian may be expected to know how to get around this type of retrieval problem, but this is not the case with end users, who are increasingly independent in their bibliographic searches.

### Research Objectives

The goal of the first stage of this research was to examine the extent of the retrieval problems caused by erroneous initial article detection at the retrieval stage in library catalogs. Consequently, two specific objectives were defined:

- Identify which initial articles have the potential to cause the most problems due to interference with non-article homographs
- Estimate the proportion (i.e., number of records with affected titles divided by total number of monographic records in the database) of bibliographic records (monographs) that are affected because of these non-article homograph words at the beginning of the title field.

The goal of the second stage of the project was to study the extent of the above-mentioned retrieval problems from the point of view of the user. To achieve this, four other specific objectives were defined:

- Determine whether end users tend to keep or omit initial articles from titles in their browse queries
- Identify which search mode is used by end users when they search the title index of the library catalog, when the titles they look for begin with an article
- Verify whether the success rate (i.e., the proportion of
retrieved records) when searching in the title index is affected by the presence of a non-article word, which is homographic to an article.

- Establish whether or not the identified problem (homographic confusion between a non-article initial word in a title and an initial article) affects the efficiency level (time and effort required to perform a search task) in title-based retrieval.

If these objectives could be carried out, it would be possible to empirically measure the extent of the retrieval problems identified. During preparation of this project, the authors noted that literature on this subject is scant; this paper aims to study this phenomenon in greater depth. Title searching is still one of the most frequent types of search in library catalogs. Making it as efficient as possible is, therefore, advisable. Broadbent’s failure analysis study revealed that around 40 percent of her survey participants came to the library looking for known items (either author or title search). Larson’s study on OPAC use also showed that, during his data collection phase (1986) in a specific catalog, the number of known-item searches (author and title) exceeded topical searches. More specifically, in 1987 Kaske measured that more than 27.5 percent of searches in a specific catalog were title searches. Matsushita’s analysis of the OPAC log at the Kunitachi College of Music Library, Tokyo (Japan) in 2000 also revealed that the most frequently used access keys are names and titles.

Research Method

The research was carried out in two phases. The first phase of the study analyzed more than 6,000 bibliographic records from the Atrium catalog (Université de Montréal). For the second phase of the study, a controlled experimental method to collect data was adopted, which made measuring the extent of the problem in one specific catalog (the University of Toronto catalog) possible. The means at the authors’ disposal being limited, this study only explored one specific catalog and prepared for a more comprehensive study of several different catalogs.

Phase 1: Case Study

For the first part of this study, the decision was made to focus on Atrium, the Université de Montréal Library catalog, as a case study. Research was further limited to monographic titles by selecting entries found in the following MARC 21 fields: 240, 245, 246, 700, 710, 711, 730, and 740, thus excluding series titles. Time and money constraints made excluding them from the sample necessary.

To meet the first two objectives, the following research questions were formulated:

**Question 1:** Which of the articles on Atrium’s exclusion list have the most entries beginning with that string of letters when not used as an article?

**Question 2:** What proportion (i.e., number of records with affected titles divided by total number of monographic records in the database) of records is affected by the deficient retrieval algorithm in Atrium?

Data collection began by identifying the 41 articles in the exclusion list used by Atrium’s initial article detection algorithm. The list is reproduced in table 1.

This list was developed locally for internal purposes and for the needs of the collection. It represents only a fraction of all initial articles listed in Annex E of the Règles de catalogage anglo-Américaines. The local list was used since research could only be performed on the articles already in the exclusion list. It should be noted that, due to system limitations, investigating the French article “l’” was not possible. This resulted in a total of 40 articles under investigation.

To answer the authors’ first research question, each article was searched individually in browse-title mode. The title index was then systematically and thoroughly scanned in order to find all the entries beginning with a non-article word homographic to an initial article. This was done by typing the article in the search box. It
should be explained here that if only the article is included in the search string (and no other words), the system does not strip the “article” and positions the user at the beginning of the title index for that word. This is how it was possible to thoroughly scan the index for each article. For each entry thus identified, the corresponding MARC record was examined to find out which field contained the problematic word. Problematic entries were recorded in a spreadsheet; those entries that resulted from the inevitable miscoding of the non-filing indicators were not retained. While examining the MARC record, the record number in field 001 was also recorded to allow the total number of affected records to be determined. This number was less than the number of problematic entries found in the index, because any given record could contain more than one title and therefore generate two or more problematic entries in the index.

To provide an answer to the authors’ second research question, the total number of affected records (those that generate at least one problematic entry in the title index) was compiled and compared to the total number of monographic records contained in Atrium at the time of research (summer 2004). It was then possible to obtain this data from the Université de Montréal Library systems office.

**Phase 2: Search Behavior Analysis**

For the second part of the study, a controlled experiment involving real users was prepared. Given the exploratory nature of this paper, and the limited means at the authors’ disposal, the decision was made to use Atrium, the University of Toronto catalog, as a case study. This catalog was chosen because the retrieval module integrates a detection algorithm designed to detect the presence of the three English articles: “a,” “an” and “the,” and offers a search interface on which it is possible, at the first level, to select a specific search mode (browse or keywords). Atrium automatically defaults to keyword searches and this is the reason why it could not be used for this part of the study. Some transaction logs of queries entered into Atrium nonetheless were used, along with the data collected from the University of Toronto search sessions, to provide additional data for question 3.

To meet the four objectives defined for this part of the study, the following four research questions were formulated:

**Question 3:** Do users usually keep the initial articles in their queries when searching the title index in browse mode or do they leave them out?

**Question 4:** When users search for known titles, which mode do they usually use: “browse” or “keywords”?

**Question 5:** What is the proportion (number of problematic records found divided by total number of problematic records searched) of monographic titles containing a word wrongly processed as an initial article by the detection algorithm that is actually retrieved by the end users, and is this proportion the same for titles not affected by this problem?

**Question 6:** Are monographic titles containing a non-article word homograph to an initial article usually harder to retrieve than other titles, in terms of time and effort?

To answer the first of these four research questions, user behavior when searching the title index of a library catalog was analyzed. The transaction logs provided by the systems office of the Université de Montréal libraries were initially examined for the searches in browse mode in the title index of the Atrium catalog for the duration of one month (October 2005). With these data in hand, checking whether users usually keep the initial articles in their title queries, or whether they leave them out, was possible.

To answer the three remaining research questions, the authors first compiled all titles containing a word that might be erroneously considered as an initial article in the University of Toronto catalog. The exclusion list used at the University of Toronto catalog consists of only the three English articles. The authors built a file of all titles that might be difficult to retrieve—in other words, the documents whose title begins with the word “a,” “an,” or “the” when this word is not an article (for example: À bout portant; An der Wegscheide; Thé ou café, Monsieur le Ministre?)—and obtained 4,384 such document titles. In order to create the data sample, only those titles were kept that were in French or in English (i.e., 1,545 titles), because participants in the study were only fluent in these two languages.

From this set of problematic titles, 24 lists were prepared, with 30 titles in each, all titles being selected at random. In order not to influence the search behavior of the participants in the study, the authors mixed different types of titles in each list. Each list of 30 titles was made up of 3 groups of titles as follows:

**Group 1:** Five titles beginning with an “ordinary” word, i.e., neither an article, nor homographic to an article (for example, Out after dark).

<table>
<thead>
<tr>
<th>Table 1. Articles in Atrium’s exclusion list</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
</tr>
<tr>
<td>ai</td>
</tr>
<tr>
<td>al</td>
</tr>
<tr>
<td>an</td>
</tr>
<tr>
<td>das</td>
</tr>
</tbody>
</table>

Note: “ai” is a contracted article in Italian. As such, it should not in theory be included in this list.
**Group 2:** Ten titles beginning with a real article (for example, *A very profitable war*).

**Group 3:** Fifteen “problematic” titles, i.e., beginning with a non-article word homograph to an initial article (for example, *À la plage*).

Throughout the rest of this paper, the first two groups of titles are usually referred to as the “non-problematic” titles and the third group of titles as the “problematic” titles.

All titles included in the lists were cataloged in the University of Toronto Library catalog, and were therefore, in principle, retrievable. The order of presentation of the titles in the lists to be searched was determined randomly, and it was modified for each list to minimize the learning factor. An example of such a list appears in the appendix.

Once these lists were prepared, 24 students at the pre-university level (first or second year of Cégep [Collèges d'enseignement général et professionnel]), enrolled in the pre-university profile (in Québec, Cégep is a required step between high school and university), were asked to try and locate the bibliographic records for the titles listed on one list. Each participant received a different list so that there would be no contamination effect. The main reason for selecting college students was to have a rather homogenous group from the point of view of exposure and experience with bibliographic searching in catalogs. Each student was requested to search all titles on his or her list using one or the other of the two search options containing or starting with as shown in figure 3.

At the start of each session, the two search options were alternated, selecting the containing option initially for one half of the participants, and the starting with option for the other half, to avoid a bias in favor of either of the two search modes, at least at the beginning of the search process. The participants were completely free to use either of the two modes at any time during the search session. The title index was preselected and the participants were not allowed to change it. Each of the search sessions was recorded using Camtasia, a software application designed to record all the operations performed on screen and to create a video that reproduces the search sessions faithfully.

Once they retrieved a record, the participants had to write down the call number on the form (see appendix), which made it possible to easily ascertain the success rate. Their answers were double-checked by replaying each video. The following information was also recorded for each title:

- Starting time: the moment when the user executes his query by clicking on the Search button
- End time: the moment when the user displays the right record (if found)
- Search mode for each query: containing (keyword mode) or starting with (browse mode)
- Number of results: in the case of keyword searches, the number of results retrieved
- Initial article inclusion or omission in string search, for titles beginning with an article

**Observations and Analysis**

For the first phase of the study, data collection was performed between July 5 and August 6, 2004. The authors were able to identify 6,360 problematic entries in the title index and believe the results would have been higher if series titles had been included because these titles often contain initial articles.

**Question 1**

*Which of the articles on Atrium’s exclusion list have the most entries beginning with that string of letters when not used as an article?*

Table 2 shows the total number of affected entries for each surveyed article and their origin in the record. A rapid survey of the data in table 2 clearly shows that some of the articles potentially were much more problematic than others. Almost half of the articles in the exclusion list never generated any problematic entries. Conversely, the article “a” alone generated 4,230 problematic entries in the index (66 percent of the total). The main explanation is that the article “a” is very common in English (and in other languages as well) and it is also a very frequently used preposition in French. For example, 1,205 documents with a title beginning with “À propos de . . .” (“All about . . .”) were noted. All these entries were the cause of retrieval problems in browse mode searches. It was also noted that problems occurred when the title began with the initial of a first name beginning as “A” (*A.B.C. contre Poirot*), and also for many acronyms (A.A.C.R., A.B.B.), or the many works beginning with A.B.C. (*A.B.C. de la lecture*, for example). The high proportion of French language works in the Atrium catalog, as compared to catalogs in...
English-speaking institutions, magnified the problem in this case. For instance, the University of Toronto catalog (being much larger than Atrium) has a lesser proportion of problematic records, quite probably because the proportion of French language resources in the former is lower than in the latter.

**Question 2**

What proportion (i.e., number of records with affected titles divided by total number of monographic records in the database) of records is affected by the deficient retrieval algorithm in Atrium?

The data presented in table 2 indicates that the total number of problematic entries in the title index was 6,360. While matching this data with the record numbers collected while doing data collection, these entries were found to be coming from 5,111 distinct bibliographic records in the catalog. Again, one should remember that any one record can contribute more than one entry in the title index. For instance, a record might have two problematic titles, one in field 245 and one in field 740.

The total number of monographic records in Atrium at the time of the data collection was estimated to be approximately 1,318,000. It may, therefore, be estimated that the proportion of monographic records affected by the initial article detection algorithm was slightly less than 0.4 percent (table 3). This proportion concerns only those titles found in six MARC fields, and this number would probably be higher if series titles had been considered.

For the second phase of the authors’ research, a log of Atrium browse-title queries was captured for the month of October 2005. For the part involving participants, data were collected at Collège de Maisonneuve (Montréal, Canada), between January 30 and February 6, 2006. Recruiting was done through posters explaining the tasks to be performed, the estimated time required (roughly 45 minutes), and the remuneration offered ($20).

**Question 3**

Do users usually keep the initial articles in their queries when searching the title index in browse mode or do they leave them out?

Analysis of the queries collected in the transaction log of the Atrium catalog indicated that users seemed to retain the initial article in their queries in approximately two cases out of three (table 4). Out of the 12,216 queries recorded in the transaction log, 1,468 queries (approximately 12 percent) were queries made to search works whose titles began with an article. This was estimated to the best of the authors’ knowledge by examining each query on a case-by-case basis, but it was not always possible to be 100 percent certain whether the title of the resource sought by the end user actually began with an article. From these queries, it was observed that the initial article was omitted in only 36.8 percent of all cases, leading the authors to believe that end users usually would rather leave the initial articles in their queries. Comparing these data with other catalogs where

<table>
<thead>
<tr>
<th>Article</th>
<th>MARC field*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>245</td>
</tr>
<tr>
<td>a</td>
<td>3,519</td>
</tr>
<tr>
<td>e</td>
<td>391</td>
</tr>
<tr>
<td>i</td>
<td>377</td>
</tr>
<tr>
<td>la</td>
<td>185</td>
</tr>
<tr>
<td>le</td>
<td>222</td>
</tr>
<tr>
<td>an</td>
<td>121</td>
</tr>
<tr>
<td>al</td>
<td>78</td>
</tr>
<tr>
<td>el</td>
<td>56</td>
</tr>
<tr>
<td>ai</td>
<td>33</td>
</tr>
<tr>
<td>los</td>
<td>40</td>
</tr>
<tr>
<td>un</td>
<td>28</td>
</tr>
<tr>
<td>um</td>
<td>24</td>
</tr>
<tr>
<td>las</td>
<td>11</td>
</tr>
<tr>
<td>li</td>
<td>17</td>
</tr>
<tr>
<td>the</td>
<td>5</td>
</tr>
<tr>
<td>bet</td>
<td>1</td>
</tr>
<tr>
<td>uno</td>
<td>4</td>
</tr>
<tr>
<td>der</td>
<td>1</td>
</tr>
<tr>
<td>ein</td>
<td>0</td>
</tr>
<tr>
<td>eis</td>
<td>1</td>
</tr>
<tr>
<td>les</td>
<td>1</td>
</tr>
<tr>
<td>19 others</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>5,360</td>
</tr>
</tbody>
</table>

*No problematic entries were found from fields 710 and 711.

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percent of total records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monographic records in Atrium</td>
<td>1,318,000*</td>
<td></td>
</tr>
<tr>
<td>Problematic entries in the title index</td>
<td>6,360</td>
<td></td>
</tr>
<tr>
<td>Affected records</td>
<td>5,111</td>
<td>0.3888</td>
</tr>
</tbody>
</table>

*Number is approximate
this feature is not present would be interesting, because Atrium users may have learned over time, from using the catalog, that they do not need to pay attention to the initial article. Nonetheless, in their study of known-item queries in OPACs, Kan and Poo noted the same behavior observed in this study.14

Similar proportions were observed when the video-recorded search sessions in the University of Toronto catalog were analyzed (see table 4). Out of 54 queries made in browse mode to find titles with an initial article, 37 queries (68.5 percent) contained the initial article, while the user had not included the article in the other 17 queries (31.5 percent). The authors’ analysis revealed that the queries where the initial article was omitted were more successful. Of 37 queries in which the initial article was retained, only 18 (48.6 percent) successfully retrieved the desired record. This rises to 88.2 percent when the initial article was removed from the query.

**Question 4**

**When users search for known titles, which mode do they usually use: browse or keywords?**

The compilation of the total number of queries made by the 24 participants to find their 30 titles indicates that more than three quarters of the queries were issued using the keywords mode (see table 5). This proportion rises to 80.2 percent if only the first query is counted for each title. Following these observations, one might assume that the users’ preferred mode is the keywords mode, but it must be remembered that the title samples submitted to the participants consisted of 50 percent problematic titles, which is not at all representative of the proportion of problematic titles in a catalog (less than 0.4 percent, according to the authors’ previous analysis). Because of the initial article automatic detection algorithm, retrieving these titles in browse mode is nearly impossible. The authors’ data reveal that none of the 360 problematic titles (15 titles on each of the 24 lists) could be retrieved using the browse mode. Analysis for all titles reveals that the last query—the query that successfully retrieved the record—was made in keywords mode in 9 times out of 10 (table 5). It is not surprising, therefore, that users ended up choosing this mode most of the time.

A chronological analysis of the queries indicates that at the beginning of the session, users were using the browse mode more often. Seventeen out of the 24 participants (71 percent) used this mode to issue their very first query, in spite of the authors taking care to preselect the keywords mode as the starting selection for half of them. In figure 4, that behavior can be seen at the beginning of the session. For the first 5 titles, both modes scored approximately the same—they were equally used. As the session continued, users progressively abandoned the browse mode for the keywords mode (only 2 percent of the queries in browse mode for the last 5 titles searched) in spite of the fact that

<table>
<thead>
<tr>
<th>Table 4. Analysis of user searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data source</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Atrium transaction log (Oct. 2005)*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Video-recorded search sessions in Univ. of Toronto catalog (Jan. 30–Feb. 6, 2006)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Transaction logs do not report query success

<table>
<thead>
<tr>
<th>Table 5. Analysis of searching modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browse mode</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Total queries</td>
</tr>
<tr>
<td>First query issued for each titles</td>
</tr>
<tr>
<td>Last query issued for each title found</td>
</tr>
</tbody>
</table>
the browse mode is known to be more efficient for locating a known document. Affirming that users prefer the keywords mode is difficult, because the overrepresentation of problematic titles in the sample gave the participants the misleading impression that the browse index mode was less efficient. A Web catalog analysis by Halcoussis and colleagues revealed that when asked to rate the level of satisfaction regarding organization of a Web catalog based on a variety of criteria, “browse-title” ranked as one of the highest search types, with a coefficient estimate of 0.919 (subject searches being set to zero as a control category), while “keywords-in-title” ranked as the lowest search type, with a coefficient estimate of -1.711.

Question 5
What is the proportion (number of problematic records found divided by total number of problematic records searched) of monographic titles containing a word wrongly processed as an initial article by the detection algorithm that is actually retrieved by the end users, and is this proportion the same for titles not affected by this problem?

The search for a known document (known-item search) for which the end user has the exact title is one of the easiest imaginable tasks in any catalog. The success rate should be near 100 percent. This is what was observed for all the titles in the samples that were not problematic (with articles and without articles combined). However, for the titles considered problematic because of the presence at the beginning of the field of a non-article homograph to an article, 2 titles out of 15 were not retrieved on average (see table 6). A t test comparison of the averages obtained reveals that the differences observed are significant (p < .0005). The authors have, therefore, concluded that titles that are considered problematic because of the presence of a word erroneously treated as an initial article by the detection algorithm are more difficult to retrieve.

Question 6
Are monographic titles containing a non-article word homograph to an initial article usually harder to retrieve than other titles, in terms of time and effort?

The time measured was from the moment the user pressed a key to launch his query and the moment the record displayed on screen. The time for keying-in the query was not counted, since titles can vary in length. System response time was noted to be minimal at all times; the time measured here corresponded mainly to the time it took for the user to recognize the correct record and select it. Titles that were not found were excluded from the average.

Analysis of the time necessary to find the records reveals that problematic titles have taken much more time on average (see table 7). Finding the titles containing an initial article took more time, compared to those without such an article, but the statistical analysis reveals that this difference is not significant (p = .062). Statistical analysis of problematic titles compared with the other two title groups combined shows that the differences observed are, in this case, meaningful (p < .0005).

In this study, in addition to time, two measurements were used to represent the effort invested by the participants to locate a title: the mean number of queries used and the mean size of the retrieved sets (for the queries issued in keywords mode) were measured (see table 7). On average, more queries were necessary to find the titles containing an initial article than to find those that did not, but the statistical analysis shows that the difference is non-significant (p = .489). Conversely, the statistical analysis comparing problematic titles with titles of the two other groups combined shows that the differences are significant (p < .0005).
On average, the users retrieved slightly larger sets (offering less precision) to find titles containing an initial article than to find those without an article. The statistical analysis, however, reveals that the difference is non-significant ($p = .763$) (see table 7). Nonetheless, the statistical analysis of the problematic titles compared with each of the 2 other title groups (with and without article) shows that the differences are significant in this case ($p < .005$ and $p < .011$ respectively). These two measurements, therefore, indicate that (on average) more time and more effort (number of queries and size of sets to browse) were necessary to locate a problematic title.

### Conclusions and Future Research

This exploratory study has supplied empirical data that are valuable if there is to be a better understanding of the phenomena of title retrieval with regard to initial articles in automated information retrieval systems. While preparing for this project, the authors' review of existing literature revealed that, while the problem is well documented on the data representation side, it is seldom examined on the retrieval side. Title searches are still one of the most, if not the most, common search type in library catalogs. It is, therefore, desirable that they be made more effective and more efficient. The results of this study show that applying an initial article detection algorithm to queries negatively affects only a small proportion of records (less than 0.4 percent of all bibliographic records in Atrium). This proportion may seem so small as to be negligible, but in reality it is

<table>
<thead>
<tr>
<th>Mean time (in seconds) to find a title</th>
<th>Mean number of queries per title</th>
<th>Mean size of the sets per title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titles without an article ($N = 5$)</td>
<td>Average 5.58 St. dev. 6.28</td>
<td>Average 1.18 St. dev. 0.37</td>
</tr>
<tr>
<td>Titles with an initial article ($N = 10$)</td>
<td>9.32 St. dev. 5.99</td>
<td>1.25 St. dev. 0.28</td>
</tr>
<tr>
<td>Problematic titles ($N = 15$)</td>
<td>19.76 St. dev. 10.14</td>
<td>1.66 St. dev. 0.33</td>
</tr>
</tbody>
</table>

Archival Products
p/u LRTS 51n2 p. 145
some 5,000 records that are thus less visible when a browse search is performed in the title index. This is not a negligible number if the acquisition and processing costs of these items are considered.

Out of the 40 articles that were examined in this study, the English article “a” is responsible for two thirds of the problems encountered. It seems that a large proportion of the problems could be solved by merely removing this article from the exclusion list. Moreover, eliminating the exclusion list altogether would eliminate the retrieval problems from the start. One may argue, however, that completely eliminating the exclusion list might introduce other problems in the searches: specifically if users inadvertently or unknowingly include the initial articles in their queries when doing a title search. The authors’ log analysis revealed that, in browse searches, only one third of the queries for titles that start with a definite or an indefinite article did not contain an article. It was observed that in about 2 cases out of 3, users kept the initial article in their query, even when these articles were ignored in the indexing process. At the time of printed catalogs (index cards, for instance), removing initial articles was mandatory in order to locate a title at the right place. End users no longer seem instinctively to remove the initial articles from their queries. In this computerized world many queries likely are generated by using the cut and paste function, which may partially explain why initial articles are retained in the queries. End users also seem to believe that keeping or omitting the initial article will have no effect on retrieval because that is the case for most of the general search engines on the Web. Using automatic detection algorithms could, therefore, be regarded as a way to adapt to the changing search behaviors of end users. Regrettfully these algorithms, as shown in this study, are not terribly sophisticated, and have major caveats, especially in multilingual environments.

The results of this study indicate that applying an exclusion list has a negative effect on a small but not negligible proportion of records, from the point of view of their visibility in the title index. The authors have observed that the success rate in finding these titles is significantly lower than the success rate in finding the other titles, since the problematic titles cannot be retrieved using the browse mode. The keywords mode is a good substitute in many cases, but retrieval may become tricky or simply impossible for problematic titles. Therefore, the authors recommend that an alternative method be developed to eliminate this problem. One possible option is to initiate some form of interaction with the end user. For example, following a search on “ti=UN resolution 435” the system could provide a feedback such as “Do you want to search un resolution 435 or resolution 435?” instead of keeping the whole procedure completely invisible.

This research could be extended to other catalogs in the future or to other environments, or be used to measure the impact on the user in a real research situation. The results of this study can be used for developing better retrieval algorithms in order to improve title searching in multilingual information systems. Since library catalogs are the entry point to many document collections, configuring the systems to maximize retrieval efficiency and success rate and, therefore, to improve customer satisfaction is essential.

The authors advocate against using detection algorithms based solely on exclusion lists since, in many cases, these mechanisms appear detrimental to end users title searches. It is preferable to include clear and highly visible instructions in the search interface, instructing end users to omit the initial article in their search. Regrettably, users are often not adequately trained or properly instructed for information retrieval in library catalogs. Before computerized catalogs existed, it was assumed that users knew that they had to remove the initial articles to find a title. Why should it be different today? It is a simple rule to learn. An alternate solution to using exclusion lists would be to ease the filing rules and allow a title containing an initial article to be filed under the article and also under the first significant word. This option, for entries starting with “The,” is recognized as a “win-win” solution by Browne. This indexing method, suggested by Nielsen and Pyle in 1995 and again, more recently, by Corrado, is already applied in some library catalogs. However, to be completely efficient, it would require recording initial articles in all MARC fields where they appear, including fields 130, 240, 246, 247, 700/710/711 (subfield t) and 730. The implementation of the non-filing control characters within the data, as proposed in Discussion Paper 2002-DP05, would certainly make this possible. The double entry solution may bulk up the title index a little but this technique makes the use of initial article detection algorithms unnecessary, because finding the titles either way (with or without the article in the query) becomes possible.

Because there are apparent advantages and disadvantages of using initial article detection algorithms, the dilemma between keeping and eliminating the exclusion list remains. Either the exclusion list is kept, allowing the cor-
rect redirection of queries containing an initial article, or it is eliminated to avoid losing track of titles beginning with a non-article word that is homographic to an article in the list. The authors hope the empirical data provided in this paper will help system designers and managers make better decisions regarding the use of such features in their catalogs.

References


15. Kan and Poo, “Detecting and Supporting Known Item Queries.”


Appendix. Example of One of the 24 Lists Given to Participants for the Retrieval Task

Liste 06

<table>
<thead>
<tr>
<th>Titre à chercher</th>
<th>Cote</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Always a loser</td>
<td>PR2987.9.A745 A75 1981 ✓</td>
</tr>
<tr>
<td>2. The most agreeable vice</td>
<td>luxlay.H593 M65 1988 pam ✓</td>
</tr>
<tr>
<td>3. A very profitable war</td>
<td>P22814 .A417 D4719 1994 ✓</td>
</tr>
<tr>
<td>4. The night after Christmas</td>
<td>JUV F16 5847mi ✓</td>
</tr>
<tr>
<td>5. Jamais contente</td>
<td>PS5907 .ML 73 ✓</td>
</tr>
<tr>
<td>6. À Vancouver sur le pouce : récit de voyage d'un étudiant à travers l'Amérique</td>
<td>HC.D911LLV ✓</td>
</tr>
<tr>
<td>7. À rire et à mourir : récits, paraboles et chansons du lointain pays, croquis, crocs, pointes très sèches, échos de la grande mort, cris et scies hors d'haleine</td>
<td>P22816 .H512 AC 1983 ✓</td>
</tr>
<tr>
<td>8. À la découverte de Shakespeare</td>
<td>PR2947 .D9 L38 T1 ✓</td>
</tr>
<tr>
<td>10. Sur le chemin Craig</td>
<td>PS9511 .L74 S87 1982 ✓</td>
</tr>
<tr>
<td>12. Les plus beaux de nos jours</td>
<td>P22807 .R83 P5 1944 ✓</td>
</tr>
<tr>
<td>Titre à chercher</td>
<td>Cote</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>15. Blessings of the table : mealtime prayers throughout the year</td>
<td>BV6235.G7 D39 1994</td>
</tr>
<tr>
<td>17. À la redécouverte de Patrice Emery Lumumba</td>
<td>DT634.L83 A5 1996</td>
</tr>
<tr>
<td>19. À la recherche de légitimités chrétiennes : représentations de l'espace et du temps dans l'Espagne médiévale, IXᵉ-XIIᵉ siècle</td>
<td>BR1024.A255 L2005</td>
</tr>
<tr>
<td>20. La mer au large : roman</td>
<td>PA670.0745 M47 1987</td>
</tr>
<tr>
<td>21. À travers chants : études musicales, adorations boutades et critiques</td>
<td>ML60 B48</td>
</tr>
<tr>
<td>22. À travers le verre du moyen âge à la renaissance</td>
<td>NK5910.M67 ROMU</td>
</tr>
<tr>
<td>23. À outrance : de cent à zero et le contraire pour flûte et viola</td>
<td>ML691.P34 L2000</td>
</tr>
<tr>
<td>24. Les femmes antillaises</td>
<td>HG1525.B4</td>
</tr>
<tr>
<td>27. A Istanbul et en Cappadoce</td>
<td>DR718.W3 1986</td>
</tr>
<tr>
<td>29. A B C on French Canada</td>
<td>CAP 00714</td>
</tr>
<tr>
<td>30. An apple a day : a holistic health primer</td>
<td>R735.B67</td>
</tr>
</tbody>
</table>
Notes on Operations

Electronic Resources Communications Management

A Strategy for Success

By Celeste Feather

Communications in the workflow of electronic resources (e-resources) acquisitions and management are complex and numerous. The work of acquiring and managing e-resources is hampered by the lack of best practices, standards, and adequate personal information management software. The related communications reflect these inadequacies. An e-resource management communications analysis at The Ohio State University Libraries revealed the underlying structure of the communication network and areas that could be improved in terms of efficiency and effectiveness. E-resources management must be responsive to the high expectations of users and other library staff. Efficient management of the related communications network increases the likelihood of a productive and successful operation.

As more resources become available in digital format and their acquisition and maintenance increase in complexity, the management of these resources in academic libraries demands greater attention. In a 2005 article, Cole described the complexities that those who manage electronic resources (e-resources) face on a daily basis. The communication network related to e-resources management also is complex. As libraries face the question of how to provide more services with fewer resources, administrators often expect e-resources acquisition units to manage more resources with fewer staff than their peer print acquisition units. Communications about e-resources management therefore are key to efficient and effective processing. An informal audit of the communication network in the e-resources unit at The Ohio State University (OSU) Libraries indicated that communications can be structured to create a more efficient operation.

The “any time any place” characteristics of e-resources create high expectations for acquisitions and access. E-resources are expensive and complex to acquire and maintain. When access or availability problems arise, users clamor for information and expect timely responses. The staff of most large libraries are not certain who performs which role in an e-resources unit. Users and staff sometimes believe that an e-resource problem will be addressed more quickly if more people know about the issue and so deluge those who manage these resources with communications, mostly via e-mail. Coping with this e-mail overload and performing complex electronic multitasking reduces staff productivity. E-resources management systems are being developed to improve productivity, but effective software that relates e-resource records, e-mail, text files, and project management work is not yet available. Creating software with such functionality and establishing best practices could dramatically improve the efficiency and productivity of those who manage e-resources.

Problem Statement

At OSU Libraries, one librarian and two library staff members are directly responsible for acquiring and man-

Celeste Feather (feather.4@osu.edu) is Electronic Resources Librarian, The Ohio State University Libraries, Columbus.

This paper is based on a presentation given at the Electronic Resources and Libraries Conference, Georgia Institute of Technology, Atlanta, March 23, 2006.

Submitted April 26, 2006; returned to author for revision and a second review June 28, 2006; resubmitted Aug. 8, 2006; reviewed and accepted Oct. 29, 2006.
aging e-resources. The e-resources unit in which these individuals work is a section within the Serials and Electronic Resources Department. The e-resources unit works closely with a librarian in the Information Technology Department, who serves as a liaison to the public services staff. This information technology position manages product trials, compiles usage statistics, manages the proxy server, contributes local information to the consortial link resolver product, and provides direct end-user support and troubleshooting in the use of e-resources. The e-resources unit staff in the Serials and Electronic Resources Department process all requests for e-resource purchases and renewals. They negotiate licenses, set up access to the resources, perform copy cataloging, manage the e-resources management module of the Millennium integrated library system from Innovative Interfaces, manage the A–Z e-journal list and MARC records profile with a third party vendor system, and troubleshoot access problems. More than half the e-resources at OSU are obtained through consortial licenses. Such heavy involvement in consortia adds complexity when the consortial resources are acquired and managed at the local level.

The e-resources unit at OSU Libraries receives and sends dozens of informative messages as part of its daily acquisition and maintenance workflow. Most of these communications are processed through e-mail, and the number of e-mail messages handled in the unit can be overwhelming for the individuals responsible. The e-mail communication is complemented by other traditional media, e.g., telephone, fax, paper mail, and in-person conversations. Timely responses are important because user expectations regarding e-resources are high and users prefer these resources because of their accessibility.

Questions arose at OSU as to whether the most appropriate types of media were being used for each type of transaction, if the communications were being processed and handled in the most efficient manner possible, and which communications should be processed in ways that would make them more accessible to a larger community. Although the communication network was not dysfunctional, improvements to maximize efficiency were needed in response to the increasing volume of work. As the work of managing e-resources evolved, the communication network needed to evolve as well.

**Literature Review**

Two fields of study, organizational communication and personal information management, are useful in gaining a broader perspective on the communications necessary to manage e-resources. Studies of organizational communication have been performed with a growing set of research methods since the 1950s. One technique, the communication audit, seeks to evaluate the effectiveness of communications systems and activities within an organization. A communication audit is a complete analysis of an organization’s communication, internal and external, that leads to a series of recommendations to upper management. These recommendations allow management to make informed decisions about improvements or directions needed in communications to achieve organizational objectives. In 1979, Goldhaber and Rogers identified the key objectives to be achieved by performing a communication audit. Communication audits are not in widespread use in the library community. Most of the library professional literature regarding communication audits emphasizes external communications and focuses on how well a library markets services and performs outreach to a user community. Cortez and Bunge introduced the notion of a communication audit for internal library communications in 1987. They noted that organizational communication is often a factor in employee stress, and that interest in organizational communication was directly related to the change and innovation then occurring.

A formal communication audit requires an objective outsider to lead the process. The study considers sociometric data and formal and informal communication within an entire organization. Portions of the research methodology also can be applied in a more focused study on a smaller segment of communication flow within an organization. Downs and Adrian provided guidelines for assessing a focused area. Among them are:

- examine how the task processes impact communication;
- determine adequacy of information exchange;
- check the directionality of information flow;
- plot communication networks;
- link internal communication to organizational networks; and
- relate communication to organizational outcomes.

Downs and Adrian also recommended guidelines for choosing methods of communication. They suggested that

- face-to-face communication is more effective for sharing knowledge;
- written communication forces clarification of complex messages;
- face-to-face communication is the best way to receive immediate feedback;
- e-mail may be best when simultaneous communication is not needed;
- persuasion works best face to face; and
- communication intended simply to inform may just as well be written.

Tourish and Hargie addressed some of the changes brought about in
the workplace by the communications revolution. E-mail in particular has served to flatten hierarchy by enabling people at all levels in an organization to communicate directly with one another without going through intermediate gatekeepers. They warned, however, that danger exists if e-mail is used so much in an organization that it displaces face-to-face communication. They also identified points to consider when auditing e-mail communications. These included the number of e-mail messages sent and received, how e-mail complements or substitutes for other means of communication, the extent to which e-mail contains information that would not be communicated by any other means, and whether goals for responsiveness have been set or are being met. Tourish and Hargie discussed information fatigue syndrome (sometimes called techno stress), describing situations in which individuals become overwhelmed by a constant barrage of electronic communications. These situations can lead to coping difficulties. Techno stress can be heightened by the expectations for high levels of service in the modern environment.

The literatures on the communication audit and personal information management are linked by the shared underlying theme of information fatigue syndrome. Hallowell labeled this neurological phenomenon attention deficit trait (ADT). ADT is caused by brain overload and appears in individuals employed in jobs that involve constant communication and constant demands for time and attention. Symptoms include decreased productivity, increased mistakes, difficulty with organization and prioritization, and the inability to focus. ADT symptoms increase gradually and usually manifest themselves in a series of minor emergencies as an individual is trying to keep up with the workload. One of Hallowell’s recommendations for addressing ADT is putting employees in an environment that promotes both face-to-face interaction and electronic communication.

Personal information management, the second field of study relevant to this research project, is a challenging area in which experts admit that no adequate software solutions are yet available. E-mail is usually at the center of the discussion because it serves so many different purposes. E-mail was developed to be a communication tool, but it also has become an archive, a project management tool, and a collaboration tool. E-mail alone is not an effective management tool. A complete integrated communications management system should include, at a minimum, e-mail, a calendar, a contacts list, a project management tool, and the embedded capability to link to other data files. Whittaker, Bellotti, and Moody noted an absence of research about what e-mail really is and what it really does within an organization. What is clear is that e-mail is being used for more purposes than those for which it was designed.

Bellotti and colleagues found that the primary reason for e-mail overload is not the quantity, but its use for task management and collaboration. They noted that current e-mail systems are inadequate for this type of work. When e-mail is used for tasks that cannot be done without the input of others, then a tracking system must be created since the threads of the conversation often are interleaved among other conversational threads in an e-mail inbox. Tracking a number of incomplete projects or tasks that have related communications interleaved in an inbox or folder results in increased stress and continuing e-mail overload. E-mail inboxes are simply not sufficient to handle this complexity of use. Bellotti and colleagues are developing a tool that would be embedded as an integral part of an e-mail system to assist in task and project management.

Venolia and Neustaedter proposed a visualization model for e-mail conversations that would enable a user to view at a glance all parts of a conversation and their relationship to each other within a hierarchy. A user could quickly see the chronology of the messages and the tree of reply relationships. Such a tool would greatly assist the tracking of asynchronous conversations.

Based on evidence that personal information management currently is poorly supported by technology, Boardman, Spence, and Sasse designed a prototype tool that would mirror and synchronize folder structures in three different areas: documents, bookmarks, and e-mail. They believe that many information management problems encountered by users are due to the fragmented nature and poor integration of the tools used. During their study, Boardman, Spence, and Sasse were surprised by the strong reactions of users toward their personal information management problems. Feelings of guilt about being disorganized and untidy, stress, and lack of control were common, and productivity suffered.

The previously discussed research is highly relevant to the management of e-resources, which requires numerous communications that currently are transmitted primarily by e-mail. E-mail often is used as a task or project management tool in this work, and the difficulties of interleaved conversations housed in an inbox that relate to documents and records stored elsewhere present additional challenges to an already complex workflow. Search features of an e-mail system are used heavily to locate relevant and related e-mail messages stored in large archival folders because no easy way to store associated messages elsewhere is readily available. The methodology of communication audits lends itself to the study of e-resources management communications because it reveals the larger network of communication relationships, directionality, and effectiveness. An objective consideration of the network of communications can iden-
tify areas for improvement, areas that cause particular stress on the individuals performing the work, and strategies that work well. A clear understanding of the communications network also enables a manager to respond more effectively as needs arise for workflow adjustment. Finally, library administrators need to be aware of the triggers for stress and overload inherent in the work of e-resources management in a complex environment. These triggers come both from the nature of the work and the inadequacy of current software tools to handle the information efficiently. This emerging specialized area of library work presents new challenges, among them those of constantly performing tasks in a highly complex communication network.

**Research Method**

The author analyzed e-resource management-related communications to and from the OSU Libraries’ e-resources unit staff during January and February of 2006. The intent was to discover how information was transmitted, if certain methods were preferred for certain types of content, who was sending and receiving the communications, and whether the communications were organized in ways that promoted productivity, efficiency, and the achievement of organizational goals. For the purposes of this study a communication was defined as an act to transmit information. The communications were classified by the characteristics of the information conveyed, including general type of content, directionality, and method used to transmit. E-mail was identified as the predominant method used for communications, and the need for closer examination of the content and number of e-mail messages quickly became clear. For two weeks in late February 2006, the e-resources unit staff members kept detailed records of all e-mail communications related to managing e-resources. Some e-mail messages were received by more than one individual in the unit, and those were recorded multiple times. The intent of the exercise was to capture the volume of e-mail workflow rather than the number of unique communications. The staff did not record other types of workplace or professional communications such as general announcements, policy discussions, local library issue discussions, and meeting announcements. Also in late February, as the final step in the audit, the author interviewed two staff members in the e-resources unit, two librarians outside the unit whose positions required them to communicate with the unit frequently about e-resource management workflow, and two librarian subject specialists who were frequent users of the unit’s services in the previous six months. The interviews elicited information about why the individuals chose to communicate about e-resources in the manner that they did, what positive and negative experiences they were having during the communication process, and what suggestions they had for improvement.

**Findings**

E-mail, telephone, fax, printed mail, in-person conversations, notes in online records, and printed documents were the methods used to transmit communications during the study. All methods except e-mail were used to transmit very limited types of content. Individuals used the telephone to transmit highly complex explanations and urgent pleas for assistance. Fax was the choice for transmitting renewal forms and license documents under negotiation whenever e-mail was not convenient. Printed mail served as the method for transmitting official copies of license documents and invoices for a small number of providers. One-to-one in-person conversations with individuals outside the unit were rare. These occurred only when an unusual or complex matter arose and the staff member outside the unit chose to speak in person rather than by phone. The communications that unit staff recorded to online records were highly specific to each e-resource involved. Unit staff members transmitted copies of printed invoices, licenses, and supporting documentation to file folders to facilitate information retrieval at a later date.

Table 1 shows the number and type of e-resource management e-mail communications recorded by unit staff members during the two-week period in February. The time to handle each type of transaction required by the e-mail varied widely. Maintenance e-mail regarding previously acquired e-resources that was sent to the e-resources unit staff presented tasks that required from a few minutes to many hours to handle, depending on the nature of the problem with each resource. Some tasks were completed with one effort, and others required multiple efforts in blocks of time spread over several days. All of the new resources requested were free. February was not an active month for adding purchased resources at OSU, and no purchase requests arrived during the two-week period that required negotiations and a long time to complete. Automatically generated invoices and alerts generally required less than fifteen minutes to handle, depending on vendor requirements and the nature of the alerts. General awareness and discussion communications from e-mail lists during this period required only time to read the messages.

The three unit staff members received 69 percent (374 messages) of the e-mail communications examined. They sent 31 percent (168) of the e-mail communications examined. The imbalance between received e-mail and sent e-mail was one indicator of the potential for stress and information
fatigue. All of the e-mail during this two-week period came from electronic discussion lists, other library staff, vendors, publishers, and automatic messaging systems. The e-mail sent by the unit staff was sent to other library staff, vendors, and publishers. No opportunities arose to communicate directly with library users during this time period, largely due to the Libraries’ organizational structure and assigned responsibilities of the unit staff.

High expectations of service caused the e-resources staff to constantly copy each other on e-mail messages just in case something might need to be addressed while one individual was away even for a few hours. An additional reason for frequently sending copies of e-mail messages to many individuals was an attempt to compensate for the demise of formal communication channels between supervisor and supervisee in the hurried workflow. Employees sometimes used e-mail to communicate with others in close proximity because it was quicker than initiating an in-person conversation, or they did not want to interrupt a colleague’s concentration or workflow.

Analysis and Discussion

The author used the Downs and Adrian guidelines mentioned previously to analyze the focused communications within the e-resource management unit. Four major categories of communications became apparent as the analysis progressed. The author named these categories darts, lobs, shadows, and spotlights, with directionality implied in their names.

Darts are the types of communications that arrive in the e-resources unit and contain all of the information necessary to perform and complete a task. Darts tend to be preformatted or automatically generated e-mail messages, but sometimes arrive from individuals with specific instructions about a task that needs to be performed. Examples of darts are messages generated by an electronic resources management system (ERMS); contents of online forms sent from other library staff who request a resource purchase, report an access problem, or request that a free resource be added to the collection; and messages sent from vendors and publishers to a group e-mail account monitored by the e-resources unit staff. The group account receives invoices, service change notifications, and other important official notices. The e-resources unit staff do not need to respond to a dart with another communication. They simply need to perform a task.

Lobs are communications that bounce back and forth between individuals in order to accomplish a task, inform, or make a decision. They arrive in the form of e-mail sent directly to individuals, telephone calls, in-person encounters, voice mail, faxes, or paper mail. Discussions on consortial e-mail lists and discussions during group meetings generally are classified as lobs. Other examples are communications among library staff about the availability of resources, the status of order requests, and the access setup for new resources. Lobs often require considerable time to handle, as each message or item needs special attention and presents a unique case. E-mail is the primary method of transmission for lobs, and the difficulties with interleaved topics of conversation presented in an e-mail inbox add to the complexity of managing this type of communication.

Shadow communications occur and are stored only within the confines of the e-resources unit. This category includes the acts of filing paper documents, storing digital files in a unit file directory, archiving e-mail, entering information in protected online record fields that are only visible to those in the unit, and conversing informally with other unit staff members. Shadow communications transmit a wide variety of content. At OSU, license documents, invoices, and information about the history of acquiring specific e-resources are stored in filing cabinets. Negotiations with vendors and agents regarding access and licenses that begin as lobs ultimately are stored as shadow communications to personal e-mail archives. Informal conversation, which in many ways is the communication channel that maintains the team work spirit and cohesiveness of the unit, often spreads knowledge about resources and operations that is never recorded outside human memory.
Spotlights, one-way communications from the unit staff to the world outside the e-resources management unit, mainly are transmitted to and stored within the library catalog and the ERMS. Access to retrieve this information may be set at different levels, such as public access to view certain records and staff access to view underlying and related records within the ERMS or the library’s integrated system. Other internal notices to staff such as those about the availability of newly acquired e-resources also are communication spotlights on the work of the unit, but the catalog and the ERMS provide the most enduring and broadest view into the work of the e-resources staff.

A streamlined and efficient communication network encourages the use of darts, minimizes the use of lobs, examines shadows to make certain that useful information is included in spotlights, and encourages the regular review of spotlights by all library staff. The complexity of the network is immediately apparent in this type of analysis. Appropriate use of each category also leads to greater satisfaction for all library staff.

All categories of communications are necessary for the successful performance of an e-resources unit. Organization of communications into the appropriate categories can increase staff efficiency and productivity. Since lobs require the most time and attention from the staff, one important goal is to examine whether some lobs can or should be transformed into darts. If certain types of communications arrive frequently with incomplete information, such as an order request without a designated fund code or an access problem without the correct title of the problematic e-resource, forms may need to be designed or redesigned to require the person completing them to enter information into specific fields. Online forms are generally very useful if they are easily accessible and create a succinct dart communication.

If vendors send invoices by paper mail that needs to be sorted and filed, they could be asked to send e-mailed invoices. Staff who place telephone calls about resource access problems could be encouraged to use online forms to report their difficulties. This ensures that the e-resources staff has the correct information with which to address the problem, rather than working from a hastily jotted note on a piece of paper after retrieving a voice mail message with incomplete information.

Shadow communications are shadows for various reasons. Some information such as database administrative login information should be communicated only within the e-resources management group. Paper is still the format of choice for some official files, such as signed license documents and invoices. Many shadows would be more useful as spotlights. Information about the status of a license negotiation that is readily accessible to all library staff could promote understanding of the process within the staff and reduce the number of inquiries the e-resources staff receive. Personal e-mail archives, which exist because transforming those communications into another format is too difficult, often contain a wealth of background information and transaction history that could be extremely useful and valuable if shared and viewed in a spotlight communication tool. Software does not yet exist that would enable an e-mail negotiation or discussion (lobs) to be linked to an ERMS record in order to provide background information for future use. Cutting and pasting is not an acceptable solution because it is too laborious. Some shadow communications become shadows because of current electronic communications software limitations. Informal face-to-face communications within the unit, as important as they are, should be monitored to make certain that key pieces of information transmitted verbally are also recorded in a way that makes them accessible in the future.

Spotlights are critical to the success of any e-resources management unit. Often useful information about e-resources is not accessible to most library staff due to inadequate management software. Information regarding the negotiation process, access rights, usage restrictions, payment history, and much more should be readily available to a large number of library staff. Accessible information helps to dissolve the mystery surrounding the management of e-resources that exists in many libraries. The work of e-resources management must be seen as integral and mainstream rather than unusual. Improving communications about e-resources management can assist libraries and their staff members in making that transition.

**Recommendations**

The analysis of the OSU e-resources management communications network revealed several ways in which processes could be improved. The improvements mentioned below are specific to OSU, but similar improvements probably could be made in many other libraries. While online forms designed to turn communications into darts were already available, they needed to be revised to update and improve the information required and transmitted to the e-resources unit. The forms needed to be renamed and links to them needed to be in more logical places. The existing lengthy names and acronyms by which they were referenced were confusing and their purposes were not always clear.

The e-resources unit staff had established a group e-mail account to receive invoices, other non-advertising important messages from vendors, and system-generated alerts from the ERMS. Over time the original purpose of the account was weakened as others
joined the group and used it for different purposes, such as receiving tables of contents from electronic journal alert services. In order to gain efficiency, unit staff took steps to return to the original purpose of the account so that communications sent to it could be trusted to be darts. Group e-mail accounts work well to raise the level of awareness of issues among the participants if responsibilities regarding workload are clearly defined and trust exists among colleagues that the appropriate person will do the appropriate work to respond to the communication. Otherwise, significant time can be lost in duplicate efforts, double-checking the work of another, and conversations to clarify who is doing what. The danger of only using personal e-mail addresses for these sorts of official communications is that if one person is absent and receives a message, no one else will be able to respond to it in a timely manner.

Since e-resources management is still new, some library staff members felt compelled to copy all individuals in the unit on all communications. While this raised the awareness of everyone in the unit about every single problem that occurred or question that needed to be addressed, the practice added to the e-mail overload that each individual dealt with on a daily basis. If a print journal issue needed to be claimed, generally one or at most two people received alerts. If access to an electronic journal ceased, often three or four people received alerts. E-resources management has evolved to the point where the matter of troubleshooting an access problem does not need to be shared with so many individuals unless it is major or unusual. For those who work with e-resources daily, an access problem with an e-journal is no more unusual than a print journal issue that needs to be claimed. A shift and change in attitude over time with encouragement by managers and administrators will likely ease this situation as e-resources integrate themselves into the daily life of all library staff members.

Another issue that arose during the course of this analysis was the need to develop more formal ways (darts) of alerting staff outside the e-resources unit when work needed to be performed, such as cataloging resources or notifying other library staff of the addition of a new resource to the collection. Notification sent in a dart communication is often more efficient since the sender does not have to worry about pleasantries and full sentence structure that would be preferred in a lob e-mail message. Also, the person on the receiving end knows exactly what to expect and what needs to be done upon receipt without having to spend time to discerning the intent of the message.

A closer examination of the communications workflow for the requests to acquire e-resources revealed a number of areas for improvement. A senior administrator for collections was required to approve every request for the acquisition of a product in electronic format, regardless of the cost. In some cases when an electronic journal was requested as an add-on to a print subscription, the cost was very low. An order for a print monograph that cost so little would not have needed approval. The workflow was established a number of years ago when every e-resource required special handling. That approach was no longer necessary in the current environment. By taking the senior administrator out of the regular workflow for every e-resource acquisition request, e-mail traffic was reduced, resources were acquired more quickly, and many fewer interleaved lob e-mail messages resulted before the final dart order request was sent. The department head of Serials and Electronic Resources also no longer felt the need to be copied on every electronic order request and problem report, so e-mail chatter was even further reduced.

The e-resources unit staff needed to make decisions about where to store certain types of information in spotlight communications since the ERMS provided the library with more places to record valuable information. Some of this information previously had been stored in order records in the library's integrated system. The ERMS will become the primary means of dissemination of information regarding e-resources management, but staff-wide access to view the records is a recent phenomenon. Training was necessary to introduce library staff to the concept of seeking information in this way. The hope is that the act of putting more and more information at the fingertips of the library staff in spotlights will reduce the number of lob messages transmitted to the e-resources unit.

During the analysis, an indication that a communication process could be improved often appeared when a style of communication did not fit into one of the four major categories. For example, when the group e-mail account established for vendor and ERMS communications could not be placed in the dart category with total comfort because a significant amount of lob traffic was sent to the account as well, that was a sign that some restructuring could improve that small area. Using e-mail filters to sort out darts so that they can be identified easily and set apart from lobs is an efficient approach. This enables workflow to be more structured and productive, and reduces the amount of time spent multitasking and dealing with interleaved conversations and messages in an e-mail inbox.

The analysis also indicated a clear need to increase face-to-face communication within the e-resources management unit in order to relieve information fatigue. Staff members began to seek opportunities to conduct business in person rather than by e-mail. Group awareness of the special factors for stress inherent in e-resources management helped to increase work-related conversations.
**Conclusion**

The audit and analysis of the e-resources management communication network at OSU Libraries revealed a need to structure the communications and to be aware of the characteristics of each type of communication in order to use them appropriately. The communications network was improved by updating and improving online request forms, reducing the number of individuals involved in certain workflow communications, reducing the number of inappropriate messages sent to an e-resources unit group e-mail account, spreading awareness among other staff about the e-mail clutter caused by notifying too many individuals of a problem, and encouraging library-wide staff viewing of ERMS records.

The data collection, analysis, and recommendations can be applied to other libraries. As workflows evolve, the communications network will need to evolve, too. One area that needs constant attention is achieving balance between communicating with too many individuals versus too few. To whom do all of the communications go, and to whom do they really need to go? Direct communications among staff members that bypass traditional chains of command and gatekeeper structures are still seen as threatening by some and as a matter of survival by others, due to the pressure of time and quantity of work. As workplaces evolve, the stress created by changing traditional communication patterns should ease.

Library subscription agents are seeking new roles in the digital marketplace as the number of printed serials subscriptions declines. Seeking their assistance for such matters as electronic journal setup, access troubleshooting, and license negotiations might relieve some of the burden on library staff in a cost effective way. These agents also could play key roles in helping to establish best practices for e-resources management between libraries and publishers. If their assistance is considered by a library, the impact on the library’s communication network also should be taken into consideration. Will information that would be useful to other library staff become shadow communications hidden in an agent’s e-resources management service or system? How easy will transferring information from an agent’s system into a local one be? Can time-absorbing jobs be reduced by enlisting the aid of an agent? Is the timeliness of the agent’s response offset by a reduced local workload? These and many other considerations will be necessary to evaluate the appropriateness of contracting with an agent to provide e-resources management services beyond acquiring a subscription.

One area of research that would assist in structuring communications more effectively is an analysis of what publishers and vendors are experiencing and expecting as they handle the management of e-resources. The library profession needs to have a better understanding of what information publishers need in the digital age. Is it possible to develop business standards that would result in a more linear workflow in e-resources management? Should library professionals encourage the development of electronic resources management systems that support more flexible and nonlinear workflows? If the workflow were less complex, the communications network necessary to support it would be as well.

As the newness of e-resources diminishes and best practices emerge, some of the intensity and anxiety surrounding the work of managing these resources will subside. For the present, however, when the management of e-resources is seen as being so critical to the relevancy and the future of academic libraries, enormous pressure exists to perform the work with utmost efficiency and accuracy. Strategies for maintaining control over the communications network for e-resources management are key components for success in this fast-paced and rapidly changing environment.

**References**


Continued on page 228
Notes on Operations

Automated Access Level Cataloging for Internet Resources at Columbia University Libraries

By Kate Harcourt, Melanie Wacker, and Iris Wolley

The explosive growth of remote access electronic resources (e-resources) has added to the workload of libraries’ cataloging departments. In response to this challenge, librarians developed various ways of providing access to electronic collections, but few dealt with the processing of free remote access e-resources, such as electronic books, Web sites, and databases. This paper will consider the various approaches taken by cataloging agencies to process Internet resources in all formats. It will then go on to describe Columbia University Libraries’ approach to cataloging free Internet resources using a combination of selector input data, an automated form able to convert the information into MARC records, access level records, and cataloging expertise.

The New Cataloging Environment

In the 1990s, with the growing popularity of the Web, more and more individuals and corporate bodies created their own Web sites and made their publications available online in addition to, or even instead of, their print counterparts. Publishers saw a marketing opportunity and quickly began to create and publish documents in electronic format. Commercial vendors promoted online over print counterparts either by using a pricing model that made continuing print subscriptions extremely expensive, or by discontinuing the print version entirely. Users and public services librarians then clamored to see remote e-resources in libraries’ online catalogs, and technical services staff had to find ways to keep up with this new and growing workload.

This challenge is likely to increase even more in the future. On October 10, 2005, the BBC reported: “In its October survey, Netcraft [a monitoring firm] found 74.4 million Web addresses, a rise of more than 2.68 million from the September figure.” Also in October 2005, the “Six Key Challenges for Collection Development” presented at the Janus Conference outlined two goals that, if implemented, would impact e-resources cataloging immensely: the digitization of all
holdings of North American research libraries retrospectively as a national project, and the shift to purchasing electronic-only items when acquiring new publications.\(^2\) As enormous amounts of information become available online, either free or through paid subscription, librarians have to tackle the ever growing task of how to select, provide access to, and manage all of these resources.

The number of cataloged non-serial remote access e-resources in Columbia Library Information Online (CLIO), the online catalog of CUL, jumped in just one year (2004 to 2005) by 359 percent, from 45,492 to 208,680. Although this number includes purchased records as well as those cataloged in-house, it nevertheless illustrates the growing demand for bibliographic access to information in electronic form. A substantial backlog of national and international online government publications existed, and the catalogers could not begin to analyze large sets of e-book collections or databases that contained other valuable resources. Selectors requested cataloging for free Internet resources using an online request form, but the requests often took a long time to fill. Paid e-resources were given priority and other e-material, by necessity, was relegated to a time-available basis.

In 2005, an existing original cataloging position was redefined to include cataloging Internet resources. Even with this additional help, Columbia’s original cataloging department could not keep up with the demand. Another approach had to be found.

The three staff members most deeply involved in e-resource cataloging formed a Work Group with the goal of establishing a workflow that would enable them to provide timely access to new publications and to process the backlog. Searching for ideas in the library literature and on Web sites of other cataloging departments, the Work Group found that many other libraries provided an online form to request cataloging of free Internet resources.\(^3\) Generally, those forms send information via e-mail to the cataloging department. While this made it easier for selectors to submit their requests, it did not help the catalogers to keep up with them.

**Literature Review**

The problem was already apparent in 1999 when Gorman posed the question “Can we afford full cataloging?”\(^4\) Gorman acknowledged the fact that full cataloging, although preferable to other bibliographic control options, is very expensive and labor intensive. At the time, he introduced the idea of applying full cataloging to e-resources of “lasting value” and to use a less expensive option—Dublin Core (DC) for others.\(^5\)

What solutions have been applied to this problem in the cataloging world?

The revised version of the Program for Cooperative Cataloging’s (PCC) Report of the Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources, published in January 2004, five years after Gorman’s article, provides some answers.\(^6\) Even though the report states that 95 percent of libraries responding to the PCC survey did catalog this type of resource, “[it] is clearly an activity that has grown greatly over a relatively short period of time, and cataloging agencies are continuing to adjust.”\(^7\) The task force found that very few of the responding libraries used metadata schemas other than MARC, such as DC, but were planning to begin using them.

A workflow that followed Gorman’s recommendation was described by Huthwaite in her article “AACR2 and Other Metadata Standards.”\(^8\) In order to provide access to their free, non-serial remote access e-resources, the librarians of the Queensland University of Technology Library and Griffith University Library use full cataloging according to the Anglo American Cataloguing Rules, 2nd edition (AACR2) for some resources determined to be important, and a DC-based schema for all others. Short records are created by reference librarians via an online form.\(^9\) This information is then converted into brief MARC records. In this approach, personal and corporate names are only accessible by keyword searching. While filling out the form, the reference librarians flag certain resources for full cataloging following their local guidelines.

Different levels of cataloging using AACR2 and MARC, however, appear to be the most popular option among the PCC survey respondents. Many make use of full, core, and minimal level records depending on the material and the needs of their institution. In addition, in 2004/2005 the Library of Congress (LC) tested and introduced an access level record for Internet resources.\(^10\) Libraries now have four levels of cataloging from which to choose, but no consistent approach on when to apply a particular level is apparent. This is still largely determined by local priorities. York University Libraries, for example, use minimal level cataloging for component parts of large collections and for Internet resources that are free with a print subscription.\(^11\) Catalogers have the option of treating the e-resources as an added copy to the print counterpart if one is available. Everything else is being cataloged as full standard. In other organizations the level of access is determined by subject specialists.

For e-journals having an equivalent print counterpart, a CONSER policy in section 31.2.3 of the CONSER Cataloging Manual explicitly allows the options of combining the description of both versions into a single record or creating a separate record for the electronic version(s).\(^12\) CONSER propagated this guideline as an acceptable policy that can be used instead of cataloging an e-journal

---

\(^{12}\) CONSER Cataloging Manual, section 31.2.3.
separately per AACR2 and the Library of Congress Rule Interpretations (LCRI). LCRI section 1.11A and LC’s Draft Interim Guidelines for Cataloging Electronic Resources allow for applying a similar single record approach for monographs.13 The OCLC document Cataloging Electronic Resources: OCLC-MARC Coding Guidelines describes this approach for any format.14

One of the questions in a 2003 survey undertaken by the Cataloging Electronic Resources/Electronic Resource Display in OPAC Task Force of the Illinois Library Computer Systems Organization User’s Advisory Group (ILCSO) focused specifically on the choice of single versus multiple records. Chen reports: “Comments from those responding to the survey leaned toward the single record method, but the decision to use a single record or multiple (separate) records for various versions of print and electronic titles had clearly not yet been settled.”15 The 2004 Report of the Task Group to Survey PCC Libraries on Cataloging of Remote Access Electronic Resources also found a large number of libraries using the single record approach in their catalogs for at least a portion of their e-journals and monographic online resources.16

Most recently, the PCC Standing Committee on Automation Monograph Aggregator Task Group listed in its Functional Requirements for Electronic Vendor Records (FREVR) Final Report the different e-book cataloging approaches currently in use in library catalogs.17 This task group described both single and multiple record options. Separate records are being created “either describing the original e-book in the bibliographic record and referring to the original edition or describing the original edition in the bibliographic record and referring to the reproduction.”18

E-resources are also made available to patrons through Web lists. Those listings can be found on many library Web sites. Most libraries provide separate lists of e-journals, e-books, and databases, some in alphabetical order, others by subject. The respondents in the ILCSO survey were “almost universally presenting some portion of their electronic holdings on Web lists instead of, or in addition to, their catalogs.”19 The same was found to be so in the PCC survey, which reported: “Over 92 [percent] of libraries (83 of 90) provide access to remote electronic resources in ways other than cataloging on the local system. Of those, 78 [percent] (65) provide access on library [Web] sites.”20 Most of those Web listings are not maintained by catalogers. In her article “Web lists or OPACs,” Anderson remarked that “for years, libraries have provided multiple and redundant access to ‘new’ media in the form of catalog entries (prepared by technical services librarians) and separately maintained lists (prepared by public services librarians).”21

Automated Cataloging Projects

Faced with the fact that none of these options seemed to solve the problem of keeping current with the workload, enterprising librarians began to think of ways to automate at least part of the cataloging process. They also discovered ways to use one data source to create both Web lists and MARC records to avoid the duplication of work done by catalogers and public services staff. Most projects of this type focused on e-journal cataloging. Anderson describes the approach developed by the Virginia Commonwealth University (VCU) Libraries in 1999.22 Using vendor-supplied data, VCU created an e-journal database for journals in aggregator databases that was searchable on the libraries’ Web site and, at the same time, was used to automatically generate minimal-level MARC records for journals that were loaded into the catalog.

A year later, at the IUG (Innovative Users Group) 2000 Conference, Jiras of the Rochester Institute of Technology reported his library’s approach to cataloging e-journals in unstable aggregator databases.23 Rollins, reporting on the process, wrote, “In a nutshell, one creates records from vendor supplied data, imports them into the catalog, and when the information changes or is out of date, one does it again.”24

The Hong Kong Baptist University Library developed an e-journal computer program (EJCOP) to provide access to their e-journals holdings.25 This project also focused on e-journals residing in unstable aggregator databases. Vendor lists and pre-existing MARC records were combined to form a single full MARC record for each full-text journal. The program was also able to convert the MARC record into HTML in order to upload the information to the e-journal list on the library’s Web site. EJCOP also was used to facilitate record maintenance on a monthly basis.

Banush, Kurth, and Pajerek described the Cornell University Library version of automated e-journal cataloging.26 The Cornell model employs the separate record approach, not just for print and online journals, but also for different electronic versions from various aggregator databases. Very brief bibliographic records are generated using vendor-supplied title and holdings data. The computer program then adds standard MARC and locally defined fields. These records are not output to the bibliographic utilities and lack some information traditionally considered to be important, such as controlled subject access, classification, and linking fields. The authors noted, however, that their approach enabled the library to provide timely title level access to all journals hidden in aggregator databases, to use this data for maintaining their e-journal Web lists, and to perform regular maintenance.
As these examples of automated cataloging projects show, the problem of keeping pace with the cataloging of e-journals, particularly those residing in large aggregator databases, has been addressed in a variety of ways. Much less effort has focused on how to automate the processing of nonserial e-resources, such as e-books, databases, and Web sites.

In 2001, the University of Florida established a nearly fully automated workflow for cataloging e-publications residing in the Extension Digital Information Sources (EDIS) database of the Institute of Food and Agricultural Sciences (IFAS). A computer program, E-pub to MARC (E2M), was able to capture the necessary information from the electronic document itself through use of a Web crawler. A MARC converter then transcribed the data into a MARC record. Cataloging rules were followed and authority control performed. The records included summaries and contents notes, but lacked subject headings, classification, and added author entries. The MARC records were loaded into the local online catalog and into OCLC’s WorldCat. The software was written for specific publications and depended on standardized HTML coding. The automatic processing of the IFAS publications using E2M ceased when the structure of the documents changed.

The Library of Congress Bibliographic Enrichment Advisory Team (BEAT) recently introduced the Web Cataloging Assistant. The cataloger copies a specific publication’s uniform resource locator (URL) into the program, which retrieves bibliographic information directly from the resource and adds generic information. The software creates a MARC record from this data and sends it to LC’s Voyager cataloging client. Catalogers update the records manually and add subject access and other necessary information. The Web Cataloging Assistant needs, just as E2M did, a “predictable and consistent layout of the bibliographic data.” It is, therefore, primarily used for works in specific monographic series that provide such a reliable structure.

In the FREVR Final Report, the PCC Standing Committee on Automation Monograph Aggregator Task Group recommended machine-generated catalog records by vendors as a way to provide title-level access to e-books residing in large aggregator databases. While this would solve much of the problem, many other publications that are not the responsibility of any vendor or publisher are available online. These include international government and nongovernmental organizations’ reports or Web sites. Libraries need to find ways to provide access to all this information.

### E-Resources at CUL

CUL’s struggle to catalog and provide access to electronic materials mirrors experiences in libraries worldwide. In February 1995, the Cataloging Department hired an e-resources/metadata cataloger to provide full cataloging, including serial holdings, for e-resources in all formats. Catalogers and managers discovered that creating and maintaining accurate e-journal holdings data was impossible and that, even with the addition of a bibliographic assistant, the Cataloging Department was not staffed to handle the volume of new digitized titles in an expanding array of formats.

In the same year, CUL sent a cataloger to OCLC to study the feasibility of using DC for certain categories of material. After much discussion and participation in the early stages of the Cooperative Online Resource Catalog (CORC) project, managers decided little would be gained through incorporating DC into Columbia’s existing cataloging activities.

CUL next began to explore ways to obtain vendor-supplied cataloging but was discouraged by the quality and scarcity of records. In 2002, Columbia cataloging administrators and the CONSER Coordinator at LC began working with Serials Solutions to develop specifications for creating CONSER-based e-journal cataloging for journals in aggregator packages. Serials Solutions searches the CONSER database for a matching bibliographic record. When a record for the e-journal does not exist, Serials Solutions creates an e-journal record by extracting agreed-upon elements (if available) from CONSER print or microform records. When no CONSER record exists, Serials Solutions creates records based on data from Thomson Gale, Ulrich’s Periodicals Directory, Serials Solutions’ own in-house catalogs, and other sources. In this way, Serials Solutions provides customers with 100 percent coverage of titles and holdings for serial aggregations. This success encouraged CUL selectors to seek additional sources for vendor-supplied MARC records in all formats. By 2006, CUL had obtained as many MARC records as possible for paid e-journals and nonserial e-resources, including U.S. government documents.

In addition to cataloging paid resources and titles within aggregations, CUL made an attempt to catalog free Internet resources. Selectors sent notifications using an e-mail form informing the cataloging staff that a resource should be cataloged. Many of the requests came from selectors in the Area Studies Department collecting materials from Latin America, the former Soviet Union, and Southeast Asia as well as from selectors in the sciences.

An even larger volume of requests came from CUL’s government information librarian. A U.S. federal documents depository since 1882, the Libraries have subscribed to the MARCIVE service for government documents since August 1994. MARCIVE, however, does not provide MARC records.
for Web sites; thus, the Cataloging Department received requests to catalog these and address other gaps in vendor coverage, including publications from foreign governments and nongovernmental organizations. The Cataloging Department gave these latter requests lower priority than paid resources because of volume and staffing constraints. All e-resources, paid or free, were cataloged at full, PCC, or CONSER levels.

Another pressure for the Cataloging Department arose when CUL began several projects to extract metadata from MARC records for remote access e-resources in order to create specialized interfaces and e-resource lists outside of OPAC, usually by form (e.g., e-journals) or genre (e.g., reference tools and indexes). These lists are located at CUL’s E-Resources Web site at www.columbia.edu/cu/lweb/eresources. The cataloging records used in these projects require special fields and procedures, necessitating extra time and expertise on the part of the cataloger. Metadata are harvested from bibliographic records and loaded into the enterprise SQL system (IBM’s dB2) that acts as a “master metadata file,” enabling real time searching and subject browse functionality. Subject access is achieved through LC call numbers extracted from the 050 field and mapped into Columbia’s Hierarchical Interface to LC Classification (HILCC).32

After most of the libraries’ e-resources were cataloged using vendor-supplied records, and a routine workflow was developed to handle the bibliographic records used for the extraction of metadata, staff members could consider how to provide bibliographic access to those not being addressed. In addition to the free e-resource categories previously identified, access was not being provided to component parts of paid databases. Selectors in many areas demanded better access to resources buried within large databases and Web sites. In addition, when paper subscriptions to many monographic series had been canceled in 2004, staff members were not available to catalog the electronic versions selected to replace them.

Access Level Records

The Work Group investigated the possibility of adopting the access level record for remote access e-resources used at LC. In 2003, LC released an initial report recommending how bibliographic control and access for these types of resources could be accomplished.33 One recommendation was a new type of record for a subset of Internet resources, one which would be rich in fields reflecting content and access and less full in descriptive fields. The record level developed by LC is an access level record that uses AACR2 and LC Subject Headings. The content designation conforms to MARC 21.

Delsey’s report Defining an Access Level MARC/AACR Catalog Record described scope, methodology, and guidelines that help define this record level.34 Appendix A in the report provided a core data set containing user tasks and evaluations made regarding importance of use of various fields and subfields. In early 2005, Reser reported on test results of access level use.35 Of special interest in this report are the results of cataloger time spent creating full records versus access records and the number of authority records not created.

Access Level Records at Columbia

In mid-2005, the Work Group examined LC’s access level model for cataloging Internet resources. Ensuing discussions centered on the core data set and LC’s decisions for access level records contained in the revised Appendixes B and C of Delsey’s report.36 The Work Group evaluated the usefulness of fields and subfields, and discussed subject analysis, main and added entries, and classification. Each member brought years of Internet resource cataloging experience to the discussion and determined that some descriptive fields were not necessary for resource discovery, did not add to description, and sometimes provided redundant information. Among the fields not used in CUL’s access record are the 260 field, all 3xx fields, and most 5xx fields. Use of the 246 field is limited to variant titles readily available. Work Group members determined that cataloger judgment should be the most important guideline when using CUL’s access record. The record contains a basic set of fields to which other fields can be added if catalogers judge them to be of value for resource discovery. LC guidelines were crucial in supporting the group’s goal of providing access and streamlining the use of descriptive fields. Work Group members adopted many of them. Appendix A at the end of this paper provides a comparison of descriptive fields used by CUL and LC in access records.

Subjects, main, and added entries, and classification follow LC’s guidelines found in Appendix C of Delsey’s report.37 Work Group members believed that these fields enrich access to Internet resources. Full subject analysis is applied to each resource using as many subject added entries and index terms as necessary. These include 600, 610, 611, 630, 650, 651, and 653 fields. Catalogers create SACO headings if necessary. Main and added entries are used when appropriate and include 100, 110, 111, 130, 700, 710, 711, 730, and 773 fields. CUL’s access level guidelines support the creation of NACO records for those headings not under control. CUL selectors use the LC classification number contained in the bibliographic record for collection development purposes. CUL
catalogers therefore continue to provide subfield $a$ of the 050; $4$ field in access level records for Internet resources. Subfield $8b$ is used only when needed to complete the class number.

The Work Group decided not to test cataloging time between full and access level records. This was based on the assumption that the results from LC’s testing would be similar at CUL. For the same reason CUL catalogers did not time access level record cataloging for comparison with those recorded by LC.

Catalogers began to use the access level record in July 2005. Selectors continued to use the same e-mail form as before and send printouts of Web resources to inform catalogers which titles needed to be included in the online catalog. During the next few months, catalogers noticed that they spent much less time finding information regarding publication data, first iterations, what terms should be used in the 246 $8$, and other elusive descriptive information. They could concentrate on subject analysis and authority control. The backlog of printouts and e-mail forms was coming under control. The application of fewer fixed and variable data fields resulted in a more standard record for Internet resources.

**Automated Cataloging of E-Resources at CUL**

The Work Group had been interested in generating MARC records from a predefined source of information since the initial evaluation of access level records. Could a MARC record be generated automatically from some source of information about each Web resource? Toward the end of summer 2005, the group began discussing this possibility. One very important realization emerged from the discussions: the workflow involved in receiving automatically generated MARC records would need to begin outside the Cataloging Department. Identification of Web resources for inclusion in the online catalog began with the selection process. Thus, the group decided that selectors would fill out an online form with data about the resource from which a MARC record would be generated.

The process of extracting data from the form needed to involve CUL library systems staff, as well. Library systems staff could not begin their work without a clear design for the online request form. The first step, then, was to define default codes and field content for the MARC record, which would be generated from the new online request form.

**Designing the Automated Cataloging Form**

The Work Group designed a new Internet Resource Cataloging Request (IRCR) form, in consultation with the Library Systems department. Library Systems staff estimates that consultations, design, and programming took thirty-five hours of staff time. The Work Group decided to make the form as simple as possible for selectors and public service librarians while at the same time obtaining sufficient cataloging data. Terminology for the different field labels was chosen in consultation with selectors in order to avoid cataloging jargon. The IRCR form (figure 1) is located on Columbia’s secure server and selectors must authenticate by inputting their e-mail ID and password in order to access the form. The only required fields are title and URL. The selector has the option of including Alternate Titles (246), Authors (7XX), Description (520), Subject keywords (653), Part of Resource (773), and a free-text “Note to Cataloger.” Selectors do not need to “sign” their requests. Instead, a field is automatically populated with the selector’s unique University Network ID (UNI).

This field is captured during user authentication and allows the Work Group to contact the selector if there are any questions. It is also used for statistical purposes. Some selectors use their UNI as a keyword search to see what has been cataloged. After the selector submits the form, a review screen is presented (figure 2).

The selector can edit or click OK to submit. If “edit” is chosen, the selector using the online form is returned to the form populated with the data already entered so that it can be revised. The last screen seen by the selector after clicking OK is a confirmation notice that includes date, title of the resource, and an assurance that a bibliographic record will appear in the OPAC in three working days.

Practical Extraction and Reporting Language (PERL) and MARC-related PERL modules are used to generate the MARC records. A Common Gateway Interface (CGI) program written in PERL generates the form and processes the data submitted. CGI allows HTML pages to interact with programming applications. The program was developed by Gary Bertchume, Senior Library Systems Analyst at Columbia University, and is freely available upon request to the authors. Programming provides an automated, single platform, Web-based solution that allows for unpredictable selector input but guarantees output for the cataloging staff whenever a form is submitted. Completely automating this process required the use of centrally maintained Unix Web servers, programs, and scripts that could run unattended in that environment. Data input into the form are gathered in an accumulation file on the Web server each time a form is submitted. A shell script is run daily to:

- copy the day’s input to a work file and reinitialize the accumulation file for the next day’s input;
- process the work file using a locally developed program, which generates a file of MARC records using the variable data found in the work file combined with a set of specified default values. Editing is done to remove control characters (e.g., tabs or carriage returns), to trim extra spaces, and to make sure that the URL is well-formed; and
- post the file of MARC records to the secure Web server and send e-mail to cataloging staff to alert them that a new file is ready and to supply the pickup URL, which allows the cataloger to access the file. The e-mail to the catalogers includes a link to a text version of the file for preview and quality control.

Discrete files for each day's accumulation are exported to the Voyager Workfile or Import file depending on cataloger preference. The file name begins with "ircr," the file creation date, and a .bin extension. A file created December 1, 2005, thus would be named "ircr_200512010200.bin." Catalogers import the records one by one from the file into the Voyager cataloging client and edit them for final production. The automatically generated MARC records contain some fields that are machine-generated through the IRCR form, and others that are supplied by the program. The coding for the fixed fields (Leader, 008, 006, and 007)
is entirely predefined and program supplied. Fixed fields are not edited by the cataloger, with the exception of language and, for PDFs only, the publication date (figure 3). This represents a further reduction of required fixed field elements from those used in CUL’s access record. The Work Group decided to take this step to take full advantage of the automated record creation. Figure 3 shows the fixed fields as supplied by the program.

The variable fields corresponding to the IRCR form are only generated if the selector supplies data. Other variable fields are program supplied and contained in every record. Figure 4 shows an example of a MARC record before review by the cataloger.

To keep the form as simple as possible for the selector, certain compromises were made and the resulting record requires careful review in several areas. All submissions generate records in integrating resources format. Until June 2006, the records defaulted to monograph format. After the implementation of the new integrating resources Leader and 008 field at OCLC, CUL’s library systems staff quickly revised the form, proving that the new workflow would survive major changes in cataloging practice. Asking the selectors to differentiate formats did not seem realistic. If the cataloger determines the title is not an integrating resource, he or she must change the bibliographic level. Catalogers currently catalog serials to full standard. The Work Group plans to apply the access level model to serials later in 2007 when PCC and LC complete their charge to extend the model to serials.38 The selector may or may not include initial articles, so the cataloger may need to adjust the 245 field for proper filing. The general material designation “electronic resource” is automatically supplied at the end of the 245 field and sometimes needs to be moved to the correct position by the cataloger if the resource title has a subtitle. The default for author is a corporate author with name in direct order (710 2), so the cataloger must retag personal names or adjust their indicators. The summary (520) is often copied and pasted from the online resource so Unicode conversion problems sometimes occur. Figure 5 shows a completed catalog record.

### Testing and Implementation

The IRCR form and cataloging workflow were tested by Work Group members before the form was made available to selectors, in two phases of testing between late September and late November 2005. The first test, done within the Cataloging Department, was to successfully generate MARC records from the information input into IRCR forms. Work files were created overnight and Work Group members were automatically sent e-mail messages containing two URLs—one for the records that would be saved to the Voyager import file and one for the text documents containing data from the IRCR forms. This test confirmed that MARC records could be generated from the IRCR forms, so the second test was implemented.

The goals for the second test were successful generation of large daily amounts of MARC records over a long period of time and successful cataloging workflow management. Participants included the three catalogers from the Work Group and a selector, who had taken part in the initial planning of the project and who was a regular contributor of e-resources titles under the previous request procedure. During October and November 2005, the selector submitted 147 records through use of the IRCR form. Each Work Group member was responsible for cataloging Internet resource titles for one week at a time on a rotating basis. At the end of this test phase, the Work Group confirmed that large numbers of

<table>
<thead>
<tr>
<th>Leader</th>
<th>008</th>
<th>Fixed-length data elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>05</td>
<td></td>
<td>Record status = n</td>
</tr>
<tr>
<td>06</td>
<td></td>
<td>Type of record = a</td>
</tr>
<tr>
<td>07</td>
<td></td>
<td>Bibliographic level = i</td>
</tr>
<tr>
<td>08</td>
<td></td>
<td>Type of control = blank</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>Encoding level = 3</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>Cataloging Form = a</td>
</tr>
<tr>
<td>06</td>
<td></td>
<td>Pub. Status = c</td>
</tr>
<tr>
<td>07-10</td>
<td></td>
<td>Date 1 = uuuu</td>
</tr>
<tr>
<td>11-14</td>
<td></td>
<td>Date 2 = 9999</td>
</tr>
<tr>
<td>15-17</td>
<td></td>
<td>Place of publication = x</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>Frequency = u</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>Regularity = u</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Type of CR = w</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>Form of orig. item = s</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>Form of item = s</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>Org. alphabet = blank</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>Entry convention = 2</td>
</tr>
<tr>
<td>35-37</td>
<td></td>
<td>Language = eng</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>Modified record = blank</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>Cataloging source = d</td>
</tr>
<tr>
<td>00</td>
<td></td>
<td>Form of material = m</td>
</tr>
<tr>
<td>05</td>
<td></td>
<td>Target audience = blank</td>
</tr>
<tr>
<td>09</td>
<td></td>
<td>Type of file = d</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>Government publication = blank</td>
</tr>
</tbody>
</table>

Note: all other available positions are coded with fill characters

<table>
<thead>
<tr>
<th>006</th>
<th>Fixed length data elements</th>
<th>007</th>
<th>Physical description fixed field</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Form of material = m</td>
<td>00</td>
<td>Category of material = c</td>
</tr>
<tr>
<td>05</td>
<td>Target audience = blank</td>
<td>01</td>
<td>Specific material designation = r</td>
</tr>
<tr>
<td>09</td>
<td>Type of file = d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Government publication = blank</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Fixed fields supplied by the program.
records could be supported, and that management of the new cataloging workflow, including a one- to three-day turnaround time, was sustainable.

The Work Group’s next major decision was whether the extra step of searching OCLC and potentially doing some cataloging there was necessary. Columbia University Libraries use OCLC as its primary source of cataloging copy and is an OCLC National Level Enhance library. The CUL corporate culture supports creating original records in OCLC and enhancing cataloging copy when necessary. Catalogers work in either OCLC or in the local system, depending on expediency and judgment. The Work Group was aware that LC opted not to search the utilities for copy before creating their access level records and wondered whether working only in the local system would be more efficient. The Work Group decided that catalogers would continue to choose where to catalog using the same criteria used for other CUL cataloging work. Influencing this decision were surprising amounts of cataloging copy found and commitment to NACO authority work, necessitating use of OCLC for name authority record creation and review.

After evaluating the second test’s results, the Work Group decided to share the new process for submitting and cataloging Internet resources with selectors and other CUL librarians. Work Group members and the selector who was a participant in the second test presented a program on the new access level record and IRCR form at a selectors’ meeting in December 2005. The presentation covered the IRCR form and its development, selector and cataloging workflow, and basic fields of the access level record. The overall response from the selectors was positive and, within days, selectors began to use the IRCR form.

**Performance and Evaluation**

The use of the IRCR form in combination with automated cataloging has provided an answer to many of the challenges created by the explosive growth of electronic information. The CUL catalogers now have a tool to provide timely access to free Internet resources submitted by selectors for cataloging. The prescribed turnaround time is three working days, but, in most cases, the records are upgraded the next day. This has had an immense impact on the workflow of the three staff members involved in cataloging free Internet resources. Instead of trying to make time whenever possible, the processing of free Internet resources has become part of the daily routine. Previously, only paid subscription databases and electronic collections received this kind of attention. By sharing the cataloging process with the selectors and employing an automated cataloging technique, the catalogers are able to concentrate their time on the creation of subject headings, access points, and authority work.

Occasionally, selectors submit more than twenty requests a day. This reduces the time available to the affected catalogers for other tasks. Cataloging staff do not feel that other assignments have suffered, since they rotate weeks for cataloging the files of requests and help each other out when a “bottleneck” develops. If the daily workload continues to increase, the Work Group may rethink some of the workflow decisions.

The Work Group timed the original cataloging of non-serial e-resources using the IRCR form for several weeks. The average cataloging time, including authority work, was sixteen minutes per record. Another expe-
rienced cataloger processed a small sample of integrating resources and e-books as full standard MARC records without help of the electronic form. The resulting average cataloging time of 31.5 minutes substantiated the group’s belief that great time savings had been accomplished. CUL catalogers feel that these time savings of 44 percent can be attributed to the combination of four factors:

- Access level records eliminate the need of searching for hidden information, such as date and place of publication, and corporate bodies.
- The automated form saves catalogers time spent on typing.
- Selectors providing summaries and keywords simplify subject analysis.
- Reliance on cataloger’s judgment rather than on strict rules eliminates the need to agonize over decisions and provides catalogers with the freedom to add additional information when necessary.

LC catalogers involved in the LC pilot project voiced mostly favorable opinions on the creation of access records, such as “a breath of fresh air,” “provided summaries were a big benefit,” or “elimination of redundancies.”

CUL catalogers agree with all of them, and add that the automated form amplifies the advantages of access records. CULs emphasis on cataloger judgment resolves possible limitations of those records. Between October 2005 and April 2006, 836 submissions were cataloged using the new method. The Work Group decided to include component parts of licensed e-resources into the workflow as well, reasoning that since the main resource already went through the acquisition process its component parts could be considered “free” and submitted along with other free remote access e-resources. This decision presented CUL catalogers with a tool to provide access to valuable resources previously hidden within large aggregator databases.

In July 2006, the Original Cataloging Department was able to report a 24.6 percent jump in cataloging production for the 2005/06 fiscal year. Cataloging managers attributed most of this increase to use of the ICRR form in combination with access level records.

One of the most rewarding outcomes of the project has been collaborative problem solving. Selectors often provide summaries, keywords, and added entries that they consider to be important. They also provide references to related print resources or suggest subject headings to related catalogers. Good communication between catalogers and selectors has become critical. The introduction of the ICRR form not only brought free Internet resources to the fore in cataloging, but also generated discussions in public services. The improved information exchange made it obvious to the catalogers and selectors that various problems arose repeatedly during the cataloging process, but were settled on a case-by-case basis. The CUL government information librarian, in consultation with other selectors and the Work Group, drafted a long-needed policy defining selection criteria for free Internet resources. For instance, free and paid content are occasionally offered on the same site. The staff members involved in drafting the policy decided that these resources are cataloged only if they make that distinction obvious to the patron. Many resources require the user to register, usually by providing an e-mail address. The Work Group was concerned that

---

Figure 5. Completed catalog record

---
some sites might pose privacy and security problems, depending on the information requested. The policy now states that CUL should continue to provide access to this type of material if considered to be of great value. The Work Group agreed to include the registration requirements in a note (506 MARC field) in the catalog record to alleviate the privacy and security concerns. This “Restrictions on Access” note displays prominently in the OPAC.

The new workflow for free Internet resources has been a great success from the technical service point of view, but does it work for the selectors? In order to answer this question, the Work Group formulated a short survey and, in early March 2006, sent it to thirty-six selectors (see appendix B). Hard copies of the survey were also distributed in the selector area of the acquisitions department. Before the deadline of two weeks, nine selectors responded. Since, by this time, not all the selectors had chosen to select free Internet resources for addition to the catalog, the Work Group decided that the nine responses were sufficient to evaluate the use of the IRCR form.

The feedback was positive. Only one respondent preferred sending e-mail messages directly to the Cataloging Department. Four selectors had used the form, while five had not but were planning on doing so. The impact on their work was generally judged as positive. One respondent wrote, “I love it. It is such an efficient way to get the record into our OPAC. Without this, I would need to baby-sit each title through the process . . .” Another selector remarked that the new form and automated cataloging process “reduce paperwork, make tracking easier, and result in faster cataloging.” The only criticism was a first impression that filling out the form might be a little more work for the selector compared to the previous submission process.

Five selectors judged the ability to track their submissions by using keyword searches and their UNI as important or very important. This feature enables them to make sure the resource was cataloged and gather their own statistics. One person stated that locating submissions by UNI is useful when handling reference questions; another used it to revisit certain sites to keep track of changes. The other four respondents either did not use this option or thought it to be useful but did not consider it to be essential.

The Work Group asked if the selectors considered the ability to contribute keywords, summaries, and other cataloging data as important. The replies ranged from “somewhat important” to “critically important.” The respondents loved being able to make use of their specialized knowledge in their subject area to point out additional titles under which a particular resource might be known, or to bring out special aspects that might not warrant a subject heading but are useful for information retrieval.

One of the replies referred to the closer working relationship between catalogers and selectors:

If I’ve already spent some time reviewing the site to determine whether it is worth adding to CLIO, then I have some knowledge of its content and that should be passed on to the catalogers so they don’t have to start from scratch. Even if they have good reason not to use my suggestions, it seems useful to suggest them. It also helps if the sites are in languages that the catalogers don’t work with. Finally, a summary may be helpful when the title of a site isn’t very informative, and increases the likelihood of discovery through CLIO keyword searches.

Only one respondent felt that this was not crucial and thought that “catalogers could handle the whole thing more efficiently and more consistently.” This selector also remarked that, in his opinion, optional selector input of keywords and summaries had not been made clear.

The Work Group asked if access level records were considered to be sufficient, both from the selector and public services points of view, or if any important information was missing. Seven respondents were completely satisfied. The other two selectors found the new model to be adequate, but also remarked that “full is better.” No respondent noted any specific data element thought to be lacking in the records.

The catalogers involved consider the feedback from the selectors as very crucial to their work. The selectors were all pleased with the one- to three-day turnaround time and found the IRCR form easy to use. Some had trouble locating it on CUL’s networked e-resources Web site. The Work Group will address this last point in the future.

Based on the responses, the new workflow appears to be as much of an improvement for the selectors as it is for the catalogers. The government document librarian, who helped the Work Group during the implementation phase of the form, commented, “As of today (Mar. 6, 2006), I have had 425 items cataloged via the Internet Resources Cataloging Request Form. In my opinion, that represents a significant addition to the electronic research material now available to Columbia University students and faculty.”

**Conclusion**

Online resources play a major role in today’s information environment. Providing access to all types of e-resource collections is crucial. CUL developed an automated cataloging workflow for free e-resources—one that includes selector input into the
cataloging process, provides online cataloging forms, and automatically generates MARC records.

In the months since the successful implementation of the IRCR form, many ideas have surfaced on how this automated cataloging workflow could be extended to other library technical services areas. The Work Group also realizes that other libraries could adapt the form and the underlying program to their own needs and projects. The form could be customized to accommodate other types of materials, such as microfilms, analytics, or to provide bibliographic access to pamphlets in vertical files. It could be adapted to handle large projects without putting strain on existing professional cataloging staff. Cataloging data also could be put into a spreadsheet instead of the form. MARC records are generated in the same way. Whether using the form or a spreadsheet, the underlying programs can be easily customized to generate resource or project specific data such as reference lists, abstracts, and notes.

Incorporation of techniques developed by the Work Group into other technical services departments and activities is a high priority for CUL. Librarians and managers are equally excited about opportunities to create quality records more easily. This new approach gives the cataloger more time to focus on subject analysis and authority control and gives patrons access to underserved areas of the collections.

References and Notes

5. Ibid., 13.
7. Ibid., 6.
18. Ibid., 2.
22. Ibid.
24. Ibid.
26. David Banush, Martin Kurnth, and Jean Fajerek, “Rehabilitating Killer


30. Ibid.


40. The guidelines for selection of free electronic resources are available upon request from the authors.

### Appendix A. LC/CUL Comparison of Variable Data Fields and Their Use

<table>
<thead>
<tr>
<th>Columbia University Libraries</th>
<th>Library of Congress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>020 International standard book number</strong></td>
<td><strong>Sa, Sz</strong></td>
</tr>
<tr>
<td><strong>022 International standard serial number</strong></td>
<td><strong>Sa, Sy, Sz</strong></td>
</tr>
<tr>
<td><strong>024 Other standard identifier</strong></td>
<td>Not used</td>
</tr>
<tr>
<td><strong>028 Publisher number</strong></td>
<td>Not used</td>
</tr>
<tr>
<td><strong>040 Cataloging source</strong></td>
<td><strong>Sa, Sc, Sd</strong></td>
</tr>
<tr>
<td><strong>042 Authentication source</strong></td>
<td>Not used</td>
</tr>
<tr>
<td><strong>050 LC Classification</strong></td>
<td>Used</td>
</tr>
<tr>
<td><strong>240 Uniform Title</strong></td>
<td>If information is readily available. Use following appropriate LCRIs</td>
</tr>
<tr>
<td><strong>245 Title and Statement of Responsibility</strong></td>
<td><strong>Sa, Sh, $b$, Sn, $p$—do not transcribe other title information ($b$) unless it provides needed information about the resource</strong></td>
</tr>
<tr>
<td><strong>246 Varying form of title</strong></td>
<td><strong>Sa, Sn, $p$; first indicator = 1; second indicator = 3 only</strong></td>
</tr>
<tr>
<td><strong>247 Former title or title variations</strong></td>
<td><strong>Sa, Sn, $p$</strong></td>
</tr>
<tr>
<td><strong>250 Edition Statement</strong></td>
<td><strong>Sa</strong></td>
</tr>
</tbody>
</table>
Appendix B. Selector Survey

1. Are you using the new Internet Resources Cataloging Request (IRCR) Form for submitting cataloging requests for free electronic resources or component parts of subscription databases? (if no, please explain)

2. Does this new form and the electronic cataloging process impact your work? If so, how?

3. a. Is the IRCR form easy to locate on SWIFT? If not, where would you expect to find it?
   b. Is the IRCR form easy to use? If not, how could it be improved?

4. You are now able to track your submissions by using keyword searches and your UNI. How important is this to you?

5. How important is it to you to be able to contribute cataloging data such as keywords or summaries as part of the cataloging process?

6. The resulting bibliographic records are less full than those for RTIs.
   a. In your opinion, is there any important bibliographic information missing?
   b. How satisfied are you with this new record model from a selector and from a reference point of view?

7. How satisfied are you with the turn-around time of 1 to 3 working days?

8. General comments?

The goal of this book is not to discuss copyright in general or even explain all of its implications for librarians and educators. Rather, the intent of Tomas A. Lipinski, who is a licensed attorney, is much more limited: “to have you read and understand the law surrounding liability and its avoidance or at least its management” (xiii). Lipinski often uses the term “risk management” as he examines in great detail the complicated legal issues surrounding copyright. He makes it very clear that libraries have choices under the current law and that there may be times when they might decide to take greater risks to achieve goals that are important to them. He wants to make sure, however, that libraries make these decisions with an understanding of these risks and of the penalties for making the “wrong” choices if they should be sued.

After the preface, Lipinski presents a seven-page “glossary of essential terms used in this book” (xxvii-xxxiii) to make sure that the reader understands the specialized copyright terminology. The text then starts with a discussion of the three types of copyright liability, proceeds to the penalties and immunities for libraries and schools, analyzes the implications of the Digital Millennium Copyright Act (DMCA), and then gives “three ways libraries and schools can limit their exposure” (301–359). Lipinski then provides three compliance tools that he considers an integral part of the main text rather than supplementary materials. Each part includes three or four chapters that follow the same format: a brief statement of the questions to be answered by the chapter, the main text, several “real-world examples,” “key points for your institution’s policy and practice,” and extensive “endnotes” that often include additional information beyond the citations. The volume concludes with a “cases” index and a “subject” index.

Lipinski makes it clear that this publication is meant to be read in its entirety and is not intended as a reference book. About midway through, he states that “[i]f the reader skipped ahead to the present discussion and remains ignorant of those concepts, now might be a good time to review those chapters” (156). He also cautions that the “summary statements may not capture the nuances of the law and are not meant as definitive statements” (xix). To gain an understanding of the legal concepts, the reader will need to work through the reasoning of each chapter, and the chapters build upon each other.

This book also destroys any belief in the certainty of the law and in the predictability of legal decisions because Lipinski makes it very clear that technology has created great uncertainty, both from the passage of new laws by Congress and from the difficulties of applying concepts that were clearer in the print environment to the digital age. He takes great care to avoid offering definitive interpretations but instead sorts through the multiple sources that judges and lawyers may take into consideration. These include the text of the laws, the legislative history, any outside commentaries, and the beginnings of case law. He concludes that it may take several major decisions by the Supreme Court before any certainty in case law emerges and that decisions in lower courts apply only in those jurisdictions, although the decisions will be used as legal arguments in other cases. He also cautions that Congress is annoyed enough by the extent of copyright infringement that new laws will most likely be passed that could add to the confusion. Finally, technology will continue to advance in ways that the original laws may not have been able to foresee, as was the case for peer-to-peer software.

Lipinski argues that Congress has been quite favorable to copyright holders in recent years by strengthening the length of copyright, shifting the burden of proof to justifying acceptable use, and passing the DMCA that criminalizes not only copyright infringement but also tools to break copyright protection even when there may be legal uses of the protected materials. He also stresses that Congress has provided exceptions for librarians and educators from the full effects of copyright liability or from the statutory monetary damages but that these exceptions come at a cost. Libraries and educational institutions must make efforts to foster copyright compliance, even to the extent of possibly requiring students to take copyright training, if these institutions are to benefit from this preferred legal status. Lipinski argues rather strongly that, while no institution can completely eliminate copyright infringement by its staff or patrons, an environment of copyright compliance reduces the risks both by minimizing the number of infringing cases and by providing the library or school with a legal defense for reduced liability.

On a more personal level, while I include a unit on copyright in my collection development course and believed that I was quite knowledgeable on the subject from reasonably extensive reading on the subject, I discovered that I was completely unaware of several important areas and, even worse, was wrong about others, most notably how the TEACH act applies to my use of copyrighted materials in
my distance education courses. Now that I can no longer claim ignorance (an appropriate factor in reducing liability both for me and for my institution) I have to make some hard decisions for next semester.

I found the book to be tough going since I have never read such a long text completely focused on legal issues. I found myself rereading sections and sought out a quiet spot free from distractions in order to concentrate. I agree in general that Lipinski has succeeded in his goal “to ensure that even the most obtuse materials presented should be accessible to the legal novice” (xix). For example, he often quotes the same section of the law multiple times as needed rather than referring the reader back to an earlier example. A few times, I had concerns about his “real world examples” where he assumed, after giving the principles, that the reader would come to the correct conclusion on whether the activity was legal or not. I would have liked him to have simply stated his conclusion. I found a few typographical errors here and there. More disconcerting was an error in the very first real world example (8) where the “employee of a public library” in the “Situation” becomes a “school media specialist” in the “Legal Analysis” a few lines below. Fortunately, my confidence in the author returned when I did not find a repetition of such errors.

In the “Foreword,” Laura N. Gasaway comments that “this should not be a reader’s first book about copyright—instead, it is an important second one” (xi). I would change this to recommend that the copyright expert in each library or educational institution read this book and that there should be such an expert if there is not. This text should also be mandatory reading for those who teach copyright. Those with a casual interest in copyright without enforcement responsibilities may find it too specialized to be worth the substantive effort involved in understanding its contents. My final comment, with which I am sure Lipinski would agree, is that this work cannot stand as the definitive tome on copyright liability for librarians and educators because new laws and new court decisions will continue to appear.—Robert P. Holley (aa3805@wayne.edu), Wayne State University, Detroit.


*Becoming a Digital Library* provides an overview of the decisions and actions, rather than a discussion of technical details or software, that culminated in the development of the digital library at Cornell University’s Mann Library. All chapters were written by digital library practitioners who represent various library departments (with the exception of systems), including public services, collection development, and technical services. Each chapter deals with an aspect of creating a digital library, such as resources, staffing, teamwork, and user feedback, which are grouped into three main categories: visions, assets, and technology.

This text is more a history of building a digital library than a guide to be consulted. Much has changed in digital libraries in terms of terminology, technology, and initiatives since it was published in 2004. The introduction states, for example, that “all of research libraries’ millions of documents will be digitized, so digital libraries must be hybrid libraries, including digital materials and pointers to other formats” (xiii). It is notable to see how far the digital library concept has evolved in the three years that have passed since this book was published. It contains a number of terms and links to resources that are dated, established and no longer considered cutting edge, or no longer available. Examples include the terms “hybrid library” and “cyberspace”; discussions of MyLibrary; the Open Archives Initiative being referred to as a new initiative (it is now a fact of life for institutional repositories); and a position description for a metadata librarian that reads more like a position for a traditional MARC-based catalog librarian with the exception that MARC and FGDC (but not MODS or METS) are mentioned. Lastly, most of the references cited at the end of each chapter are dated in the late 1990s and early 2000s.

Key concepts such as metadata and digital preservation are noted briefly. This text lacks a chapter specifically devoted to metadata, which is unfortunate since this is what drives resource discovery and retrieval. Instead, it is included in various chapters in the book. There is also no mention of the *Functional Requirements of Bibliographic Records* (FRBR), which date back to 1998 and are often included in discussions of metadata schema and applications.

Digital preservation is covered in Chapter 3, “Resources for the Digital Library,” in a section titled “Creating the Digital Library: Providing Access to Historical Material” (76). A URL is provided to a Cornell document on recommended specific requirements for depositing image collections in a central archive repository. While this document is dated 2001, much of it is still applicable to image formats and digitization.

The term “institutional repository,” which is now more commonly used than “digital library,” appears nowhere in this text, although there is a 2002 Scholarly Publishing and Academic Resources Coalition (SPARC) reference to it available on the Web.

Chapter 3 also contains a section titled “What is a Digital Library?” that provides five definitions that are no longer used. They are an interesting illustration of how far the concept of a digital library has evolved in three years.

The definitions are: (1) stand-alone digital library or SDL, (2) federated digital library or FDL, (3) harvested digital library or HDL, (4) gathered digital library or GDL, and (5) services for using the digital library or SUDL (50-52).
In contrast, a relevant working definition of “digital library” as put forth by the Digital Library Federation (dated 1998) is included: “Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities” (xii).

Despite the fact that some of the information in this text is dated, it contains many universal concepts that are applicable and provide good information, such as the chapters on personnel (specifically hiring and training), collection development policies, teamwork, and project implementation and management. This text also touches on issues that are still challenges for digital library initiatives, including copyright, staffing for the digital library, paying for the digital library, and getting appropriate support from one’s administration. Some of the chapters include sidebar descriptions of projects and experiences, often written in the first person, by project participants or leaders; these are insightful and complement the text. Although different individuals wrote the chapters, the writing flows and is cohesive. This is often not the case for works with multiple authors, and speaks to the editor’s contributions.

A quote about engaging the entire institution in digital library initiatives and mainstreaming digital projects is relevant in current context and is also indicative of the spirit of cooperation that likely existed at Mann Library: “The organization relies on the skills of catalogers and the talents of programmers to develop metadata structures, while the institution depends on the vision of public services and the knowledge of selectors to create a repository of information resources” (2). Furthermore, Chapter 2 (“Mainstreaming”) indicates that many of the skills needed to build a digital library are already present in libraries in acquisitions (purchasing, licensing), cataloging (access to resources), and public services (experience with information tools). Becoming a Digital Library illustrates how quickly terms and concepts related to digital library technology change. It provides an interesting look at the digital library development of a leader institution and provides some universal information about personnel, teamwork, and project management that are appropriate to all library environments.—Mary Beth Weber (mbfecko@rci.rutgers.edu), Rutgers University, New Brunswick, N.J.

Reference

Electronic Resources Communications Management continued from page 211


Index to Advertisers

Archival Products ...........................................................................................................................................................199

EBSCO .....................................................................................................................................................................159, 187

Library & Information Technology Association ...........................................................................................................179

Library of Congress.................................................................................................................................................cover 2

Library Technologies...............................................................................................................................................cover 3