

Library Resources & Technical Services

ISSN 0024-2527

April 2002

Volume 46, No. 2

ARTICLES

- Characteristics of Resources Represented in the OCLC CORC Database** 39

Tschera Harkness Connell and Chandra Prabha

- Technological Change and the Scholarly Communications Reform Movement** 50

Reflections on Castells and Giddens

Richard Fyffe

- Inventory at Brooklyn College, 1998–1999** 62

An Original Method

Judith W. Wild

FEATURES

- Book Reviews** 72

Margaret Rohdy, Editor

Seymour Lubetzky: Writings on the Classical Art of Cataloging, *reviewed by Martha M. Yee*

Subject Analysis in Online Catalogs, *reviewed by J. Bradford Young*

The Invisible Web: Uncovering Information Sources Search Engines Can't See, *reviewed by Gregory Wool*

Maps and Related Cartographic Materials Cataloging, Classification, and Bibliographic Control, *reviewed by Kay G. Johnson*

American Library Association
Association for Library Collections & Technical Services

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 BOARD

Editor and Chair

John Budd

University of Missouri

Members

Kathleen R. Brown, *North Carolina State University*

Lynn S. Connaway, *NetLibrary Incorporated*

Jeanne M. Drewes, *Michigan State University*

Martin M. Kurth, *Cornell University*

Jack Montgomery, *Western Kentucky University*

Linda L. Phillips, *University of Tennessee*

Richard E. Sapon-White, *Oregon State University*

Stanley J. Wilder, *University of Rochester*

Barbara A. Winters, *University of Georgia*

Ex-Officio Members

Charles Wilt

Executive Director, ALCTS

Miriam Palm, *Stanford University, emerita*

Editor, ALCTS Newsletter Online

Douglas Raber, *University of Tennessee*

Book Review Editor, *LRTS*

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, \$55 per year in U.S., Canada, and Mexico, and \$65 per year in other foreign countries. Single copies, \$15. 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: John Budd, Editor, *Library Resources & Technical Services*, University of Missouri, 221M Townsend Hall, Columbia, MO 65211; (573) 882-3258; fax: (573) 884-4944; e-mail: buddj@missouri.edu. Advertising: Bill Coffee, c/o Benson, Coffee & Associates, 1411 Peterson Ave., Park Ridge, IL 60068; (847) 692-4695; fax: (847) 692-3877. ALA Production Services: Troy D. Linker, Kevin Heubusch; Ellie Barta-Moran, Angela Hanshaw, Kristen McKulski, and Karen Sheets. 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; 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 p. 220–21 of the October 2001 issue and on the LRTS Web page at www.ala.org/alcts/lrts. Copies of books for review should be addressed to Douglas Raber, Book Review Editor, *Library Resources & Technical Services*, School of Information Science, University of Tennessee, 804 Volunteer Blvd., Knoxville, TN 37996-4330 (raber@utk.edu).

© American Library Association 2002

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.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

Publication in *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.

Characteristics of Resources Represented in the OCLC CORC Database

Tschera Harkness Connell and Chandra Prabha

More and more libraries are providing access to Web resources through OCLC's Cooperative Online Resource Catalog (CORC) and, by extension, OCLC's WorldCat database. The ability to use a database to its maximum potential depends upon understanding what a database contains and the guidelines for its construction. This study examines the characteristics of Web resources in CORC in terms of their subject matter, the source of the content, publication patterns, and the units of information chosen for representation in the database.

The majority of the 414 resources in the sample belonged to the social sciences. Academic libraries and government agencies contributed more than 90% of the records for resources in the sample. Using the Anglo-American Cataloguing Rules, 2d edition (AACR2) definitions for publication patterns that are part of the upcoming 2002 amendments reveals that nearly half of the sample fell into the category of integrating resources. Identifying units of representation of the resources described was more difficult. Existing definitions for Web units in development are not adequate to describe all of the resources in the sample. In addition, there is wide variability in the units of representation chosen for inclusion by the libraries contributing records, resulting in little predictability in what units of information might be found in the database.

One way for libraries to provide access to Web resources is simply to provide a connection to the Internet from public terminals. More and more, however, library staff are providing more than a connection. They are providing enhanced access by organizing and presenting those resources that they consider particularly useful to their users in ways that will help users find them. Some libraries are providing access through the library's Web page, using the library Web page as a portal for resources selected by traditional selection criteria. Others are providing access by including records representing Web resources in the online catalog so that users can find items covering the same subject matter, in all formats, from a single database. Many libraries are doing a combination of both.

One aid to librarians wishing to provide access to Web resources through the catalog is the Online Computer Library Center (OCLC) Cooperative Online Resource Catalog service (CORC). For end-users, CORC is a subset of OCLC's Online Union Catalog, WorldCat, which offers descriptions and holdings information for millions of resources in all physical formats. Descriptions are contributed by participating libraries. The CORC portion of the database presents bibliographic records and pathfinders representing electronic resources. The bibliographic records are descriptions of electronic resources;

Tschera Harkness Connell (connell.17@osu.edu) is Associate Professor and Serials Coordinator for the Ohio State University Libraries, Columbus. **Chandra Prabha** (chandra@oclc.org) is Senior Research Scientist with the OCLC Office of Research, Dublin, Ohio.

The authors would like to acknowledge the research support of Robert Bolander and Suhas Aswath.

Manuscript received February 14, 2002; accepted April 2, 2002.

the pathfinders are subject guides of resources on a topic. From a library processing point of view, CORC is a system for creating metadata to describe electronic resources. It also allows the metadata creator to choose from several encoding formats such as MARC, RDF, and HTML meta tags. If a record for a resource is in CORC, CORC works similarly to other OCLC input software in that the person processing the resource can copy catalog and export the record into a local system. However, if there is no record for the resource in CORC, the CORC software creates the basic record. Inputting staff must provide a URL for the resource and choose from the offered metadata formats one to be used for the description. CORC then automatically creates a basic record for the resource, using software to harvest information from the resource itself. Once the basic record is created, staff edit the record and export it to the local system.

This article reports on a part of ongoing research at OCLC. This part of the project was a joint project of OCLC and the Ohio State University Libraries. The OCLC Web Characterization Project (<http://wcp.oclc.org>) addresses basic questions about the Web—how big it is, what it contains, how it is evolving. This project examined the characteristics of Web resources that have been identified, evaluated, selected, and described by librarians in the OCLC CORC database. The specific goal of the research reported here was to determine the nature of Web resources described through CORC in terms of their publication patterns and their units of representation. The unit of representation is the level at which the library represents a chosen resource that has a hierarchical relationship to other resources. The publication pattern of a bibliographic resource refers to the completeness or projected completeness of the resource at the time it is released (that is, published). This article also examines the subject matter and source of the resources. The term “source” is used to describe the origin of the Web resource and to describe the library or information agency contributing the descriptive record for the Web resource. Our examination of source determined whether the institution creating the description of the resource was the same institution that had made the Web resource available on the Web. Resources made available and cataloged by the same institution were categorized as internal resources of the contributing institution.

Background

Publication Patterns

In cataloging, a resource that is intended to be complete in a finite number of releases has been considered monographic. Cataloging codes and practice have been less clear in defin-

ing nonmonographic resources. Monographic publications are commonly contrasted with serials on the basis that serials continue indefinitely. However, the Anglo-American Cataloguing Rules, 2d edition (AACR2), definition of a serial includes an additional dimension that is unrelated to completeness or time: a serial must be issued in successive parts (*Anglo-American Cataloguing Rules* 1998, 622). Resources that do not meet the added criterion of the serials definition (e.g., loose-leaf publications) are difficult to catalog because they are largely ignored in AACR2. In the environment of tangible formats, these types of publications are proportionately few and catalogers have developed means to work around the lack of guidelines on how to handle them. In the electronic environment, however, the number of resources that continue indefinitely but are not issued in successive parts is great. Electronic resources, although they may continue indefinitely, are also often revised continuously. And as they are revised, their form and content may evolve. Discussions about new definitions of publication patterns developed from the recognition that there is no provision in AACR2 to indicate variances from the serials model of successive parts for publications that continue indefinitely.

Hirons et al. refine definitions of publication patterns by dividing all resources into two categories, finite and continuing. Finite resources “are complete or intended to be completed” (Hirons et al. 1999). Finite resources include monographs. “Continuing resources are those that are intended to be continued for an indeterminate period (e.g., serials, updating loose-leaf publications, databases, etc.)” (Hirons et al. 1999). Building on the work of Hirons and others, the Joint Steering Committee (JSC) for the Revision of AACR provisionally approved the definition of continuing resource for addition to AACR2 noting that “[c]ontinuing resources include serials and ongoing integrating resources” (Joint Steering Committee 2001). The JSC defines an integrating resource as a “bibliographic resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole. Examples of integrating resources include updating loose-leaves and updating Web sites . . .” (Joint Steering Committee 2001). A serial is a “continuing resource issued in a succession of discrete parts, usually bearing numbering, that has no predetermined conclusion. Examples of serials include journals, magazines, electronic journals, continuing directories, annual reports, newspapers, and monographic series” (Joint Steering Committee 2001).

Units of Representation

The importance of indicating the unit of representation when describing the design of bibliographic instruments (e.g., the online catalog), has been well stated by other writ-

ers. Wilson calls for makers of bibliographic instruments to state the design specifications of each instrument so that users will be able to get maximum benefit from its use (Wilson [1968] 1978). He notes that “[t]here is a distinction between not finding what we are looking for and finding that what we are looking for is not there; the former is a failure, the latter a negative success” (Wilson, 59). Further, he states that without knowing the “specifications” for the design of the database, it is not possible for the user to make a distinction between the two (Wilson). Bates, in writing about standards for systematic bibliography (which includes the library catalog or database), states similarly that “. . . the bibliography should not only list materials, but also state information that enables the bibliography to be located *relative to the rest of the graphic universe*. In order to accomplish the latter, we must state precisely what is and is not covered in the bibliography . . .” [emphasis in the original] (Bates 1976, 13). While Wilson and Bates differ slightly in their respective lists of required specifications, both consider it essential to define the units that are represented in the database. Bates refers to these units as bibliographic units (Bates 1976, 14), whereas Wilson writes of the unit of representation or “the unit for listing and description” (Wilson [1968] 1978, 61). For the remainder of this article, we will refer to the specification for the unit as the *unit described* or the *unit of representation*.

For tangible resources, the unit described in library catalogs has been determined, in effect, by publishers’ packaging and libraries’ collection development policies. The issue of the unit of representation has never been well addressed by cataloging codes. What libraries have acquired (book, serial, video, etc.) is what has been described. Individual libraries have had the option to describe groups of books, such as those in a series, instead of each individual book, but the decision is often made on the basis of publisher presentation and/or local collection development policy. If the publisher provides an individual title and numbering for each item in the series, then the library is more likely to describe the individual items. Or, if the library plans to buy all or most of the series, the series may be cataloged as a unit, especially if there are many items in the series and the series is well known. The library may also decide to provide some additional description and access to selected parts, but again that is a local decision.

In the general introduction to AACR2 (1998), the issue of what unit to describe is sidestepped by the following statement: “The rules cover the description of, and the provision of access points for, all library materials commonly collected at the present time” (*Anglo-American Cataloguing Rules* 1998, 1). Materials “commonly collected” are the domain, the universe from which materials are selected for inclusion in the database. For these materials, the issue of what unit to represent or describe is addressed in the scope

notes of chapters devoted to a particular type of publication. The scope notes set parameters for the type of material covered by the chapter and in doing so, define the unit to be described under the rules of that particular chapter. For example, chapter 2 presents the rules for describing separately published monographic printed items (i.e., books, pamphlets, and printed sheets). The chief source of information for cataloging these items is the title page. Following the guidelines of this chapter means that units described are whole books, whole pamphlets, and entire printed sheets. Other examples of the units of library materials to be described include whole sound discs and tapes, whole movies and videos, and whole runs of serials.

AACR2 does provide a means for analysis or, “preparing a bibliographic record that describes a part or parts of an item . . .” (*Anglo-American Cataloguing Rules* 1998, 299). However, in practice, analysis is infrequently done. Cataloging a chapter in a book, a single reading from a sound disc, or the music from a motion picture requires a great amount of effort on the part of the cataloging agency. In terms of overall design of online catalogs, AACR2 and common practice for choosing whole units for representation result in a database of resources represented broadly, or stated differently, a database with low granularity in terms of the information units described.

The organizational tradition for archival material also takes a broad approach. Modern archival science is based partially on the assumption that the significance of archival materials “is heavily dependent on the context of their creation, i.e., their *provenance* . . .” (Hensen 1989, 4). The consequence of this principle is that the cataloging manual *Archives, Personal Papers, and Manuscripts* (APPM) “approaches the problems of archival cataloging principally at the collection level. . . . [To emphasize] individual components at the expense of the whole collection may tend to obscure the intrinsic importance of the whole” (Hensen 1989, 5). In the scope note for the chapter on description, the APPM provides a list of materials that a collection may contain: correspondence, memoranda, photographs, maps, drawings, pamphlets, broadsides, newspaper clippings, motion picture films, and computer files (Hensen 1989, 9).

One of the difficulties in cataloging new materials is this issue of what to represent. Although the introduction to AACR2 states that the rules can be used “as a basis for cataloging uncommonly collected materials of all kinds and *library materials yet unknown*” (*Anglo-American Cataloguing Rules* 1998, 1), consensus on the unit of representation for new materials has to evolve (emphasis added). If the new materials are not that different from other materials for which conventions have been established, then consensus may be quick to form (for example, videocassettes and CD-ROMs have parallels in film and 33 1/3 rpm sound recordings). For Web resources that are

digital versions of printed/paper documents, or serials, librarians have tended to choose the same unit of representation as they have for the print counterparts. However, Web resources such as Web sites are not mirrors of tangible resources, and the need for clear definitions has been recognized. The identification of meaningful, distinct Web bibliographic units was a fundamental issue for bibliographic control of Web resources (O'Neill and Lavoie 2000). They also suggested a framework for definitions: "Rather than corresponding to physical objects, meaningful bibliographic units on the Web are found within the structure of Web-accessible information. . . . If their use is complemented by unambiguous definitions, Web sites and Web pages represent useful concepts for identifying bibliographic units on the Web" (O'Neill and Lavoie 2000, 55). They proposed the following definitions for Web page, Web site, and Web collection based on the structure of URLs:

Web page: A distinct information unit composed of one or more HTTP-accessible files, referenced and accessed in its entirety by a single URL (O'Neill and Lavoie 2000, 57).

Web site: A collection of interlinked Web pages residing at the same Web host (59).

Web collection: A portion of a Web site, consisting of multiple Web pages, that represents a distinct resource (59).

O'Neill and Lavoie's definitions are built partially on the work of the World Wide Web Consortium (W3C). Lavoie participated in the Web characterization activities of the W3C that resulted in a 1999 working draft document, *Web Characterization Terminology and Definitions Sheet*. Although no longer an active document of the W3C, this document provides some additional practical definitions for Web resources. The W3C definitions of Web site publisher and Web subsite, in addition to the definitions of a page, a site, and collection, that were refined by O'Neill and Lavoie, have been used for the research reported here. A Web site publisher is a "[p]erson or corporate body that is the primary claimant to the rewards or benefits resulting from usage of the Web site, incurs at least part of the costs necessary to produce and distribute the site, and exercises editorial control over the finished form of the Web site and its content" (Lavoie and Nielsen 1999). A subsite is a "[c]luster of Web pages within a Web site, that is maintained by a different publisher than that of the parent Web site, or host site. The subsite publisher exercises editorial control over the Web pages comprising the subsite, perhaps restrained by some broad guidelines imposed by the host site publisher" (Lavoie and Nielsen 1999).

Method

In preparation for this study, a pilot was conducted using records randomly selected from those entered into the CORC database from October through December 1999. The principal purposes of the pilot were to develop a standard methodology for examining sites and to determine which characteristics of sites would be used as the focus of the second phase of the project reported here. Specific objectives of the pilot were to test the application of existing characterization schemes for describing distinct Web bibliographic units, and to categorize the subject content of those units, the institutional origins of the content of those units, and the institutional sources of records describing those units.

The second phase of the project involved a proportional sample of member-created records, taken over the 12 months of July 1, 1999–June 30, 2000. A sample size of 384 records ($[n = (1.96)^2(.5)^2/ (.05)^2]$) was needed for a 95% confidence level. An additional 77 records were drawn for the sample (461 total) so that NetFirst and InterCat records could be eliminated and still meet the needed sample size. A sample of 461 accounts for the possibility that 20% of records would be nonmember records: $[(384 + (384)(.20)) = 461]$. NetFirst records were eliminated because they are created by OCLC, not member libraries. InterCat records were eliminated because although they are created by OCLC member libraries, they are not created using CORC. After eliminating nonmember records and records for which no usable URL could be determined, the final usable sample was 414 records.

Resources represented by records in the sample and the records themselves were captured on a CD-ROM so that each resource could be examined as it appeared at the time the sample was drawn. In some cases, multiple screen shots of a resource were captured if the Web address accessed a page that served as a collective listing for several different resources, and the bibliographic record described a resource off that page that could not be accessed directly. All resources were then characterized by source, subject matter, publication pattern, and units of representation.

The characterizations were made by examining each resource. Records were used to assist in the identification of resources only in those cases where a URL was not enough to identify the resource selected by the library. For example, in one case, the URL was to a site that gave a collective title listing for several agricultural technical reports. Examination of the record revealed that the resource selected by the library was an individual report, not a composite site.

Because we had learned during the pilot project that characterization of the resources in terms of unit described was the most difficult determination for the project, the characterization was performed by several individuals and then discussed in groups. OCLC staff who had been involved in

the development of the definitions, as well as the current project team, characterized the resources. In group sessions discrepancies were examined and normalized, if possible, by definition refinement.

Results

Description of the Sample

Contributing Libraries, Internal and External Resources, Subject Matter

As part of the examination of the resources, data were collected on the contributing libraries to determine who was using CORC for cataloging. Academic libraries and government agencies were by far the greatest contributors of records in the sample, contributing a total of 92% of the resource descriptions. Government agencies included national, state, and city governments, governmental departments such as the U.S. Department of Agriculture, governmental regulatory commissions, the military, and law enforcement bodies. Public libraries were not considered government agencies for this study and were counted separately.

Out of 414 records, 67% (278) were contributed by academic libraries. Twenty-five percent (104) were contributed by government agencies, and of those, 23% of the total (94) were contributed by U.S. federal and state agencies. Public libraries contributed only 3% (13) of the records. All other groups (associations/foundations, corporations/business, and networks/consortia) contributed fewer than 10 records each (<3%) (see table 1).

Part of the promise of the Web has been the potential for individuals, groups, and institutions to make available resources that had never been widely available in the past. For that reason the authors were interested in determining to what extent CORC was being used by libraries and other information agencies to describe their own unique resources. In the sample, 21% (88) of the resources were characterized as internal resources and 78% (323) were characterized as external to the institution cataloging. For three resources it was not possible to make a determination.

At first glance, the portion of internal resources (21%) in the sample may seem low; but, given the amount of preparation required to make resources available electronically (e.g., digitization of the resources, database infrastructure creation, metadata assignment, and Web design), it is quite positive that one-fifth of the resources examined in this study were internal or local resources. Said another way, one-fifth of the resources in the sample were "new" resources to the general public. Prior to the Web these resources were only available by traveling to the contributing library or information agency.

Resources in the sample were classed broadly using the Library of Congress classification. The majority of the resources were classed as social sciences (see figure 1). The largest single category, in fact 14% (57/414) of the total sample, was commerce-related. Examples of commerce-related sites include company and bank Web sites, transportation and commerce regulations, and product catalogs. Other types of social science resources well represented were national, state, and local governmental Web sites. Arts and humanities sites included artifacts of history such as photographs and historic maps, reproductions of paintings, and works of literature. The sciences were represented by sites emphasizing technical issues in agriculture, science, medicine, and military/naval science. Science resources included sites devoted to a particular research project or grant, a particular disease, and even an armed forces technical training curriculum.

Publication Patterns

Using the definitions of Hirons et al. (1999) for finite and continuing resources, 42% (173/414) of the resources in the sample are finite (see table 2). Sixty-nine percent (120/173) of the finite resources mirror traditional monographic resources such as art reproductions, dissertations/theses,

Table 1. Records Contributed to the CORC Database by Library Type ($n=414$)

Contributing Libraries by Type	No. of Records	% of Records
Academic Libraries	278	67.1
Government (U.S.) Libraries	94	22.7
Public Libraries	13	3.1
Government (Non-U.S.) Libraries	10	2.4
Network/Consortia	8	2.0
Association/Foundation Libraries	6	1.4
Corporation/Business Libraries	5	1.2
Total	414	99.9

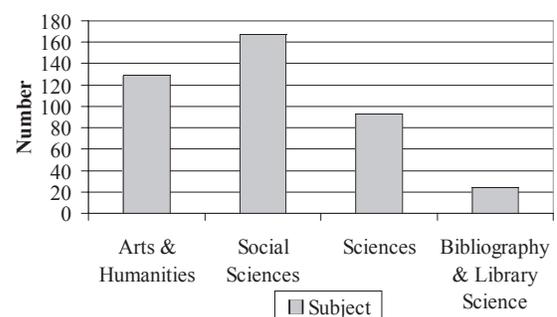


Figure 1. Subject Distribution of CORC Resources ($n=414$)

books (including exhibit catalogs), and documents. An example of a document is shown in figure 2. The other 31% (53/173) of the finite resources include individual encyclopedia entries, maps, photographs, and archival collections that appear to be complete, for example, the papers of a former university faculty member and department head. Individual photographs such as those from Northwestern University's Curtis collection of historic photographs (see figure 3) make up 25% (44/173) of the finite resources or 11% of the entire sample.

Continuing resources comprise 58% (241/414) of the sample: 80% (192/241) of these are integrating resources and 20% (49/241) are serials. Overall, serials make up 12% (49/414) of the total sample. Examples of integrating resources include the University of California, Berkeley resources on Iberia (see figure 4) and the Naval Research Laboratory, Chemistry Division home page (see figure 5). These are both integrating resources because, as they are updated, the updates become an integral part of the whole.

Table 2. Publication Patterns of Resources ($n=414$)

Publication Patterns	No.	%
Continuing Resources	241	58.2
Integrating Resources	192	46.4
Serials	49	11.8
Finite	173	41.8
Total	414	100.0

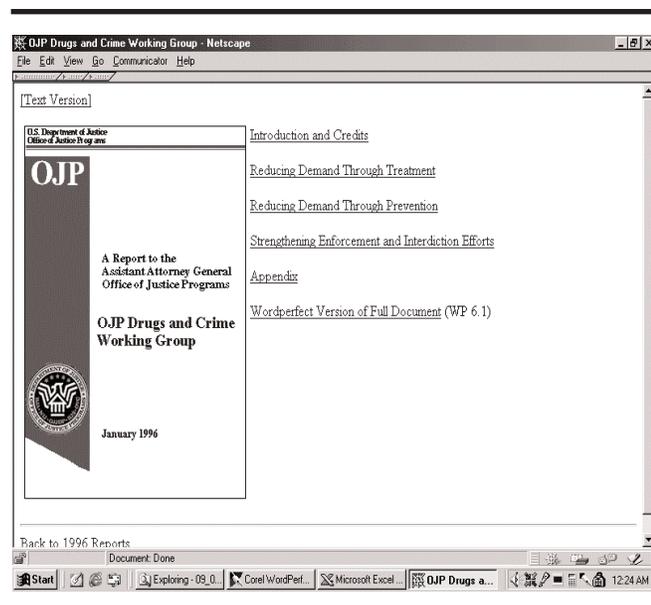


Figure 2. Example of a Monographic Web Resource (Report to the Assistant Attorney General, Office of Justice Programs OJP Drugs and Crimes Working Group, 1996)

Unless a snapshot has been archived, there is no way to view the resource as it existed before the update. Figure 6 depicts a serial (*Commission of Preservation & Access Newsletter*).

To examine units of representation, resources were categorized by two sets of definitions: (1) the traditional physical units of resources in libraries of the twentieth century, and (2) the Web structure units proposed by W3C (1999) and O'Neill and Lavoie (2000). Using physical unit definitions involved categorizing the 233 resources (or 56%) that mirrored tangible resources. First, these were categorized by the types of library materials AACR2 presents as *commonly collected*. Within these types, the resources were further broken down in terms of the unit represented, for example, book, chapter, encyclopedia, an entry from an encyclopedia, serial, a single issue of a serial, etc. The 181 resources that were primarily "loose-leaf" in nature were not categorized by the use of tangible resource comparisons.

- *Resources mirroring tangible resources.* Of the resources of the sample that could be characterized by their tangible counterparts, 63% (147/233) were whole items matching units of materials cataloged in libraries of the twentieth century (see table 3). Examples include reproductions of paintings, whole books, complete databases, dissertations, theses, newspapers, and serials. Thirty-seven percent (86/233) of the materials that mirrored tangible resources would traditionally be considered parts of units and possibly candidates for analysis. In some cases, these resources would not have been described in the catalog, but instead placed in library vertical files. Examples of analytics are 2 entries from an encyclopedia and 15 individual issues from various serials. Twenty of the resources were time-sensitive, similar to brief printed pamphlets or fact sheets—materials traditionally placed in a library vertical file.
- *Categorization of resources using Web-structure definitions.* All but two resources (which were eliminated due to technical difficulties) were characterized by the definitions for Web resources of the W3C and O'Neill and Lavoie (2000). Categorization was accomplished by two groups of individuals working independently. One group categorized all the resources using the W3C definitions for Web sites, subsites, and pages. Individuals in the second group included a category for Web collection. Results are included in table 4. Pages appeared most frequently, totaling more than one-third (35%) of all resources in the sample.

Discussion

Most of the discussion that follows centers on observations and issues relating to the publication patterns and units of representation of the resources. First, however, a few observations about the contributors of the records in the sample. It is not surprising that academic libraries contributed most of the records to CORC in this sample. The data for this study come from records contributed to CORC during FY2000. Prior to July 2000, CORC was still largely experimental and in very active development. In introducing CORC, OCLC used a very different approach than it had historically used for its products and services. CORC was released for public participation early in its design stages in order to encourage evaluation of and collaboration in its development. Many academic librarians and administrators consider it part of their mission to advance the field of library and information science through experimentation and testing of new ideas and are therefore often willing to participate in developmental projects. The sample time frame was the first year that OCLC charged for using the CORC service and promoted it as product rather than prototype. By then many academic libraries had been involved in CORC for some time. Additionally, high participation by academic libraries reflects contributions to OCLC's WorldCat database as a whole.

Publication Patterns

Data show that more than 40% of the resources in the sample were finite (table 2). Twenty-nine percent (120) fit the traditional images of finite resources, including individual works of art, books, documents, dissertations/theses, law and legislation, and reports. Ten of the resources were actual "electronic books," such as copies of published monographs. By far the largest portion, 25%, of the 173 finite resources in the sample were individual photographs.

Continuing resources made up 58% of the total sample. Most of these (80%) were integrating Web sites; 20% were serials. Overall, serials made up 12% of the entire sample. This figure is actually double the proportion (6%) of serials in WorldCat (OCLC 2001). The disproportionately high numbers of Web sites and individual photographs in the sample may indicate that CORC is being used by libraries primarily for special projects or possibly for experimentation with new types of resources. Web sites and individual photographs are not the types of resources that would easily fit into traditional library work flows. CORC provides a convenient means for trying out new software and work flows and for gaining experience with new types of resources on a project basis.

There were instances when, without knowing the contributing library's intention, categorization of the resource would have been very difficult. For example, figure 7 probably depicts a serial. If the bibliographic resource of interest



Figure 3. Example of a Monographic Web Resource (Native American Indian Photo, <http://hdl.loc.gov/loc.award/iencurt.cp10005>)

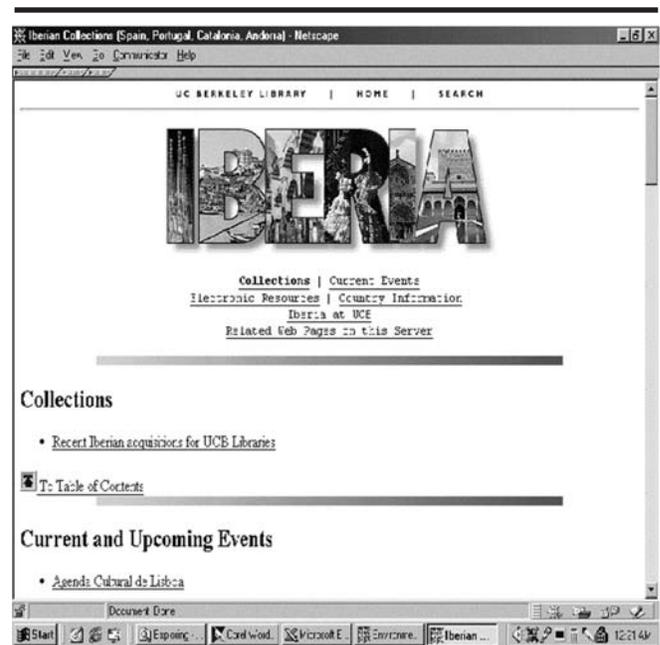


Figure 4. Example of Integrating Web Resource (Iberia, University of California at Berkeley Library, www.lib.berkeley.edu/Collections/Romance/iberia.html)

(to the cataloging agency) is the journal, *Professional Candy Buyer*, the record will represent a serial. If the resource of interest is the Web site that includes the journal as well as other resources for candy buyers, then the record will represent an integrating resource. In this case, the contributing library chose to represent *Professional Candy Buyer* as an integrating resource.

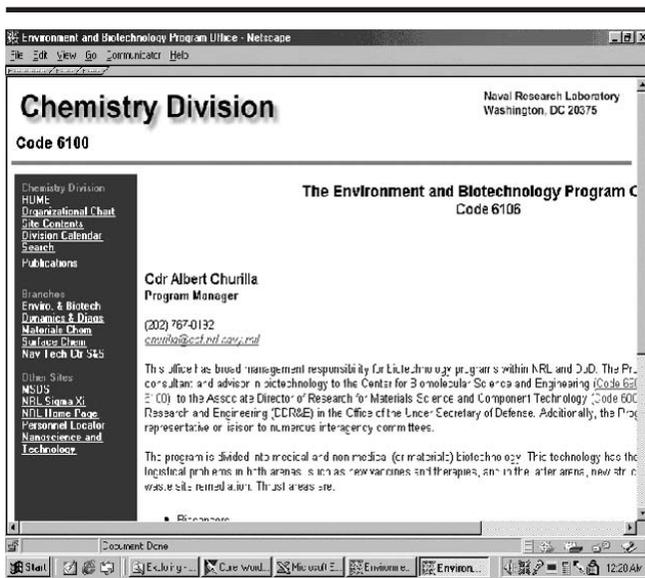


Figure 5. Example of an Integrating Web Resource (Naval Research Laboratory, Chemistry Division, Environment and Biotechnology Office Home Page, www.chemistry.nrl.navy.mil/6106/index.html)

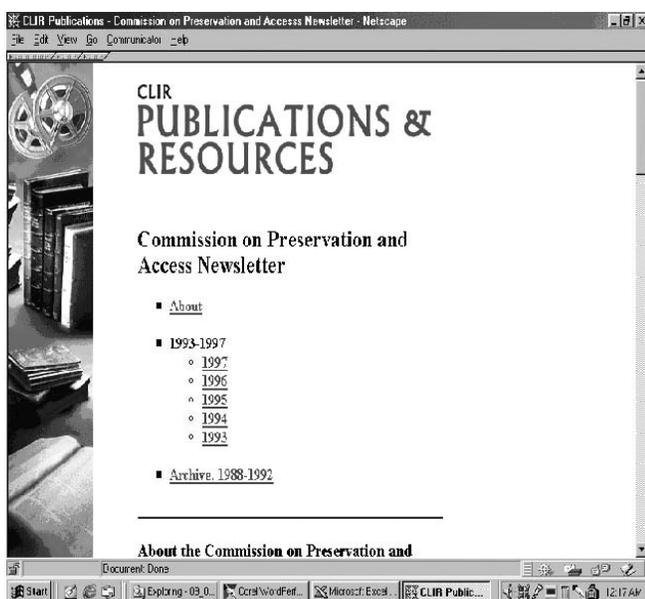


Figure 6. Example of an Integrating Serial Web Resource (CLIR Commission on Preservation and Access International Newsletter, www.clir.org/pubs/pain/pain.html)

In this study, statistics were not recorded on the frequency of ambiguous cases such as the *Professional Candy Buyer*. However, ambiguity in how to handle publication pat-

terns is not new. For example, it has long been a local choice to decide how to treat monographic series. Two libraries may choose different solutions for the same series. Any given library will implement different policies for different titles. In some cases the series will be cataloged as a unit, as a serial. In other cases, individual titles in the series will be cataloged separately as finite resources. Over the years libraries have developed guidelines for their local decisions. These guidelines include factors such as whether the library's intent is to purchase the entire series, whether the individual volumes in the series have individual titles, and how the series is treated by indexing services. Also important is how other libraries have handled the series and whether cataloging copy is available. Similar guidelines have yet to evolve for situations such as the one illustrated by the *Professional Candy Buyer*. As guidelines do develop, librarians will be able to provide a level of predictability for users of their catalogs.

Units of Representation in Traditional Terms: Web Resources That Mirror Tangible Resources

Even though more than half (223/414) of the resources in the sample mirror traditional resources, only two-thirds of these (147/223) were represented at unit levels comparable to common practices for handling their tangible counterparts (table 3). The 147 resources that were handled traditionally comprise 36% of the entire sample. Analytics and ephemera make up 21% (86/414) of the sample. Examples of ephemera included an announcement of a town meeting agenda, an advertisement for an upcoming music festival program, and an online "brochure" of free trees available as part of a promotion for Arbor Day. There were numerous instances of photographs that are clearly part of collections of photographs but that were described and represented individually. As discussed earlier, this practice is contrary to archival cataloging principles (Hensen 1989, 5), and while this option is not precluded by AACR2's chapter 8 for graphic materials, photographs in general purpose libraries have tended to be described as a group. "If the item being described consists of two or more separate physical parts . . . , treat a container that is the unifying element as the chief source of information . . ." (*Anglo-American Cataloguing Rules* 1998, 202). Groups of published photographs or slides are likely to have containers, but following the tradition of archives, even original photographs have been most commonly described as a group (collection) based on general subject matter or provenance. This is especially true when the description of the set is to be integrated into a general topic online catalog. In contrast to common practice, most of the examples of photographs in the sample have been described individually. If a unifying subject has been assigned for purposes of collocation, it has been treated as a series name.

The fact that resources we normally refer to as analytics and ephemeral documents make up 21% (86/414) of the sample means that the resource group we studied is a very different group than the resources represented in a traditional catalog. Said another way, a database for these CORC resources is a very different database in make-up than the online catalog. A fifth of the resources in this CORC sample are small units which contributes to the creation of a database with high granularity; in contrast, the typical online catalog has low granularity. Searching a database with high granularity involves a different set of expectations, search strategies, and vocabulary than does searching a database with low granularity. Patrick Wilson discusses this issue in *Two Kinds of Power*. In the chapter on reliability, he writes of evaluation of bibliographical instruments: "We cannot know how much power is made available to us by a bibliographical instrument unless we know both the plan or Specifications of the work and the quality of workmanship. And each of the separate elements of the Specifications offers a field for the evaluation of performance . . ." (Wilson [1968] 1978, 127). He considers a number of evaluative questions including, "Have the units to be separately listed been chosen correctly and consistently?" (Wilson, 127). He states that if this question (and the others he has posed) are answered affirmatively, the bibliographic instrument "can then be pronounced reliable or trustworthy. . . . The overall reliability or trustworthiness of an instrument depends on the exactness and accuracy and consistency with which the rules embodied in the Specifications are applied . . ." (Wilson, 127).

CORC records become a part of the larger OCLC WorldCat database. WorldCat, because of its birth and growth in the last third of the twentieth century, is a traditional database in terms of the units of library resources it represents. Its specifications have been largely governed by the application of AACR2 and other library standards. The results from this study seem to indicate that CORC participants, by contributing records for smaller units, are changing the traditional "specifications" of the WorldCat database. This is not a conscious redefinition of WorldCat; there is very little discussion of what units are to be represented in catalogs. Much of our practice has been formed of habit and tradition. Because WorldCat is so large, the effects of inconsistency in how units of information are represented may not be noticed to any great degree for many years. However, in time the lack of specifications for units in the database could affect users' ability to predict and to find the information they need.

Units of Web Integrating Resources: Web Sites, Subsites, and Pages

At the time this study was designed and characterization of the raw data performed, the working draft "Web

Table 3. Level of Representation of Resources That Mirror Tangible Resources, Level of Representation of Monographs and Serials (n=233)

	No. of Records	% of All Records in Sample (n=414)
Whole units	147	35.5
Art reproductions	5	
Books	10	
Collections	6	
Databases	7	
Dissertations/Theses	5	
Documents	80	
Newspapers	3	
Serials (whole)	30	
Series	1	
Analytics	86	20.8
Documents (ephemeral)	20	
Encyclopedia entries	2	
Maps	5	
Photographs (individual)	44	
Serials (single issues)	15	
Total	233	100.0

Table 4. Resource Categories As Indicated by URLs

Web Resource Categories	No. of Records	% of Records
Collections	117	28.3
Sites	85	20.5
Subsites	42	10.1
Pages	146	35.3
Variances or Undetermined	24	5.8
Total	414	100.0



Figure 7. Example of a Serial Integrating Web Resource (Professional Candy Buyer. www.retailmerchandising.net/candy/Default.asp)

Characterization Terminology and Definitions Sheet” (Lavoie and Nielsen 1999) was an active document of the W3C. Since that time, the document has been dropped as a work item with no evidence on the W3C site that it has been incorporated into another document [www.w3.org/1999/05/WCA-terms/]. A purpose of the working draft was “to bring clarity to the terms often used when talking about the Web” (Lavoie and Nielsen 1999). However, the authors certainly found it difficult to categorize resources according to the definitions.

Categorizing Web sites and Web pages was relatively easy, especially when using O’Neill and Lavoie’s (2000) clarifications for site and page. In addition to the base definition that “a site is a collection of interlinked Web pages [or a complete set of Web pages] residing at the same Web host,” O’Neill and Lavoie refined the definition by adding that the “access point for the Web site is the home page—the Web page accessed using the base URL of the Web host” (O’Neill and Lavoie 2000, 59). Similarly, O’Neill and Lavoie refined the definition of page (“a distinct information unit composed of one or more HTTP-accessible files, referenced and accessed in its entirety by a single URL” (O’Neill and Lavoie 2000, 57) by adding two practical considerations:

- A Web page consists of the set of HTTP-accessible files that are viewed simultaneously in a Web browser when the page’s URL is accessed (O’Neill and Lavoie 2000, 57).
- A Web page located at a given host can be accessed by starting at the host’s home page and traversing a sequence of links appearing only in other pages located at the same host (O’Neill and Lavoie 2000, 58).

The individuals who categorized the resources in this study had high agreement in their coding of Web sites and Web pages. Two hundred thirty-one resources (55.8%) were assigned to categories of site or page (see table 4). Of these there were only 19 for which categorization differed among those performing the categorization. This translates to 92% (231/250) agreement in categorization of sites and pages.

It was more difficult to categorize those resources that were neither site nor page. The W3C document provided two intermediary categories: subsite and collection, both dependent upon the determination of the role of the publisher of the resource. (For W3C definitions of subsite and Web site publisher, see the “Background” section of this article.) Determination of the role of a corporate entity has never been easy and, in fact, the recognition of this led to the change between the Anglo-American Cataloging Rules (1967) and the Anglo-American Cataloging Rules, 2d edi-

tion (1978) in rules for determining whether a corporate body has principal responsibility for the creation of a resource. AACR2 limits cataloger discretion and only allows assigning principal responsibility for a work to the corporate body under very narrowly defined situations. The authors of this study, in trying to determine the role of the Web site publisher, often referred to AACR2 rules for *construction* of names for corporate bodies. Decisions were made on the basis of whether the corporate body listed on the resource being analyzed as a possible subsite would be considered, according to AACR2 guidelines, as independent of or subordinate to the corporate body listed on the site. O’Neill and Lavoie’s (2000) set of definitions were somewhat easier to apply in that they did not include the category *subsite*. They used *collection* as the only intermediary category and defined it independently of the publisher’s role: “A portion of a Web site, consisting of multiple Web pages, that represents a distinct resource” (O’Neill and Lavoie 2000, 49).

Another difficulty the authors recognized in the assignment of categories is the problem of multiple addresses for a single resource. In the experience of the authors, multiple addresses for a single resource is a situation that arises frequently when describing Web resources, especially government resources. Because of the design of this research, which captured each resource in time (and place) on a static CD, multiple addresses were not a factor in categorization. However, the issue did arise as researchers re-examined Web sites online to learn more about the resource and its relationship to other sites. The possibility of multiple addresses is a weakness of definitions based on the structure of the Web address.

For the reasons discussed above, the authors were dissatisfied with categorization of resources according to definitions of site, page, and possibly collection and/or subsite. The ease with which definitions can be applied is an important consideration in terms of work flow and productivity, and these definitions were not easy to apply. However, even had they been, the question must be asked as to how meaningful these definitions would be to the general user. Traditional definitions of book, serial, film, etc. have a basis in use as well as a tactile reality. Users carry a book, a serial, or a film. Users do not carry or, we suspect, even routinely search Web resources by the definitions of collection and subsite that were used in this study. Users are likely to think of Web resources in terms of their relationships with other resources. Is the desired site *off* another? Is it a part of something else? O’Neill and Lavoie’s (2000) definitions of site and page do address this navigational aspect of relationship. Unfortunately, only 55.8% of the resources in this study could be categorized as either Web site or Web page.

Conclusions and Suggestions for Further Research

The goals of this research were to determine the nature of Web resources described through CORC in terms of their publication patterns and their units of representation. We also examined the subject matter and source of the resources. The resources in the sample covered the complete range of subjects as represented by the Library of Congress classification system. Most resources were contributed by academic libraries, reflecting the contribution patterns of libraries and information agencies to OCLC's larger union catalog, WorldCat. We found it very positive that despite the many practical and financial barriers to digitization, one-fifth of the resources in the sample are unique resources, resources owned only by the information agencies that contributed the record(s) for the resources. Prior to the advent of the Web, these resources would have been unavailable (except through travel) to the general public.

Definitions to aid in the discussion and handling of Web resources are needed so that clear specifications can be developed for databases representing these resources. The cataloging community's definitions of finite and continuing resources to describe publication patterns were clear and easy to apply to the resources in the sample. These definitions are important in that they enable the user to predict change in the resource described.

However, definitions for Web units need further development. When AACR2 was written, its rules applied to *materials commonly collected*, which at that time matched the physical packaging of tangible resources. As a result, the unit described for these materials tended to match the content of the package. While this may have been adequate in the environment of tangible materials, only half (53.6%) of the resources in this sample have a printed/tangible counterpart to aid in their recognition. The concept of *materials commonly collected*, which described the domain of the traditional catalog, is no longer practical as a substitute for clear definitions of units of representation in the catalog.

The definitions for units of representation developed by the W3C and others (Web page, Web subsite, Web collection, and Web site) were also tested, with limited success. The definitions for subsite and collection were difficult to apply and resulted in a great deal of inconsistency among the results of researchers categorizing resources for this study. The definitions for Web site and Web page were easy to apply, but were applicable for only 55.8% of the resources in the sample. Additional development of meaningful definitions is needed to build databases that provide predictability for the user. The additional definitions should be based on how users use Web resources, how they identify them, how they navigate among them, and how they

remember them for future reference. These are issues for further study.

In addition to needing unambiguous definitions for identifying units of representation, there is a need to decide what units will be represented in the database. The data from this study show wide variation in the units of Web resources described by libraries and information agencies. Traditionally, library catalogs represented resources broadly. By contrast, the CORC sample represented a large number of resources that were small units of information (photographs, individual entries from an encyclopedia, Web pages). On the other hand, the sample included large units such as archive collections, books, serials, and Web sites. This fact raises the issue of the affect of mixing the size of units described in the same database. How do units of representation affect the ability of users to predict potential outcomes of a search and thus tailor searches for maximum success? This is an important topic for further study.

Works Cited

- Anglo-American Cataloguing Rules*, 2d ed. 1998 rev. Prepared under the direction of the Joint Steering Committee for Revision of AACR; edited by Michael Gorman and Paul W. Winkler. Chicago: ALA.
- Bates, Marcia. 1976. Rigorous systematic bibliography. *RQ* 16, no. 1: 7-26.
- Hensen, Steven L., comp. 1989. *Archives, personal papers, and manuscripts: A cataloging manual for archival repositories, historical societies, and manuscript libraries*. 2d ed. Chicago: Society of American Archivists.
- Hirons, Jean, with the assistance of Regina Reynolds, Judy Kuhagen, and the CONSER AACR Review Task Force. April 1999. Revising AACR2 to accommodate seriality: Report to the Joint Steering Committee for the Revision of AACR. Accessed Dec. 16, 2001, www.nlc-bnc.ca/jsc/ser-rep0.html.
- Joint Steering Committee for the Revision of AACR. 2001. Outcomes of the meeting of the Joint Steering Committee held in Washington, DC, USA, April 2-4, 2001: Revising AACR2 to accommodate seriality. Accessed Dec. 26, 2001, www.nlc-bnc.ca/jsc/0104out.html.
- Lavoie, Brian, and Henrik Frystyk Nielsen, eds. 1999. *Web characterization terminology & definitions sheet, W3C working draft 24-May-1999*. Accessed Dec. 7, 2001, www.w3.org/1999/05/WCA-terms/01.
- OCLC Online Computer Library Center. 2001. *Annual report 2000/2001*. Dublin, Ohio: OCLC Online Computer Library Center.
- O'Neill, Edward T., and Brian L. Lavoie. 2000. Bibliographic control for the Web. *Serials Librarian* 37, no. 2: 53-69.
- Web characterization terminology and definitions sheet, W3C Working draft May 24, 1999. Brian Lavoie and Henrik Frystyk Nielsen, eds. Accessed Dec. 7, 2001, www.w3.org/1999/05/WCA-terms/01.
- Wilson, Patrick. [1968] 1978. *Two kinds of power: An essay on bibliographical control*. Reprint, Berkeley: Univ. of Calif. Pr.

Technological Change and the Scholarly Communications Reform Movement

Reflections on Castells and Giddens

Richard Fyffe

Reconceiving and reorganizing collection development practices around the evolving processes and products of the scholarly communications cycle has become one of our profession's fundamental opportunities. However, our increasing use of market mechanisms and digital technologies to rationalize the production and distribution of scholarly information poses significant risk that business cycles and the obsolescence of hardware and software will lead to the inadvertent loss of significant portions of our intellectual heritage. This article introduces a theoretical framework for understanding the relationship between academic culture and digital technology as they relate to scholarly communication and library collection development, drawing chiefly on the work of the social theorists Daniel Bell, Manuel Castells, and Anthony Giddens. The article suggests that Castells's theory of the network society and Giddens's account of disembedding, expert systems, and risk as hallmark features of modern society together point us toward a more candid recognition that the fragility of digital systems and the resulting possibility of significant cultural loss are intrinsic features of the new landscape of scholarly communications. Moreover, acknowledging this risk is an important dimension of successful reform of the scholarly publishing system.

Scholarly communication has become a guiding metaphor for academic librarianship, and reconceiving and reorganizing collection development practices around the evolving processes and products of the scholarly communications cycle has become one of our profession's fundamental opportunities (Atkinson 1996, Atkinson 2000). At the same time, however, our increasing adoption of market mechanisms and digital technologies to rationalize the production and distribution of scholarly information—while promising a resolution to the cost crisis in scholarly publishing and bringing us within view of a truly national or international scholarly collection distributed across a network of cooperating repositories—also poses significant risk that business cycles and the obsolescence of hardware and software will lead to the inadvertent loss of significant portions of our intellectual heritage.

The challenges posed by digital technologies for long-term preservation of data and cultural objects have been extensively documented and discussed (see,

Richard Fyffe (rfyffe@ku.edu) is Assistant Dean for Scholarly Communication, University of Kansas Libraries, Lawrence.

Manuscript received September 5, 2001; accepted November 26, 2001.

for example, MacLean and Davis 1998). As Donald Waters writes,

[D]igital information and the technologies on which they depend are extremely fragile. Their fragility makes it highly uncertain that digital libraries can endure over time and it causes one to wonder about the durability of their supposed benefits. Rapid cycles of change and obsolescence infect the hardware and software products now in common use to create new knowledge (Waters 1999, 193–94).

Waters continues: “The challenge of creating the deep infrastructure needed to sustain digital records of knowledge over time consists, at least in part, of marshaling a complex set of political, economic, and technological forces toward the development of a system of organizations that have come to be known generally as digital libraries” (Waters 1999, 195). That is, the solution to the challenge of assuring the continuity of digital information is not just—or even mostly—technological; rather, it is economic and political or, more broadly, cultural. For example, economic models must be created for digital objects that may be used seldom, if ever, but that still assure long-term revenues to cover the ongoing growth and replacement of hardware and software; and governance models must be developed that define rights and responsibilities, that facilitate effective decision making, and that can be perpetuated across many institutional generations.

However, less consideration has been given in our professional literature to the question of the effect of technology on the cultural conditions necessary for the preservation of digital information. Scholars such as Lewis Mumford (1934), Harold Innis (1972/1950), Marshall McLuhan (1962), and Elizabeth Eisenstein (1979, 1983) have shown in various ways that representational technologies are not culturally neutral, that the material form of information storage and transmission conditions the practices of scholarly communities. As Eisenstein writes of the new fifteenth-century technology that integrated type molds, moveable type, and the printing press: “As an agent of change, printing altered methods of data collection, storage and retrieval systems, and communications networks used by learned communities throughout Europe” (Eisenstein 1983, xiv). Standardization of copies within a printed edition, for example, made it possible “for scholars in different regions to correspond with each other about the same citation and for the same emendations and errors to be spotted by many eyes” (Eisenstein 1983, 51).

These are scholarly practices that we now take for granted but that became widespread only by virtue of a particular form of representational technology. Moreover, Eisenstein points out, one cannot treat printing “as just one among many elements in a complex causal nexus, for the

communications shift transformed the nature of the causal nexus itself. It is of special historical significance because it produced fundamental alterations in prevailing patterns of continuity and change” (Eisenstein 1983, 273; emphasis supplied). As the academy and the larger society of which it is a part make increasing use of the technologies of digital representation and networked communication, it is worth asking how traditional scholarly practices and the values those practices embody might be affected.

In this article I will introduce a theoretical framework for thinking about the relationship between academic culture and digital technology as they relate to scholarly communication and library collection development, drawing chiefly on the work of the social theorists Daniel Bell (1976), Manuel Castells (2000), and Anthony Giddens (1990). I will argue that Castells’s theory of the informational society and Giddens’s account of disembedding, expert systems, and risk as hallmark features of modern society together point us toward a clearer recognition that the fragility of digital systems and the resulting possibility of significant loss of scholarly literature in digital form are intrinsic features of the new landscape of scholarly communications. Moreover, acknowledging this risk is an important dimension of successful reform of the scholarly publishing system. Librarians must recognize, in particular, that in initiating or taking leadership for certain reform activities, they are taking this risk on behalf of the scholarly community they serve; to maintain credibility, they must be candid about the nature of those risks. In this respect, the ideal of “seamless access” to information products and services, insofar as it obscures the legal and economic complexities of the scholarly communications system, may inhibit the cultural transformation that will be required to create lasting reform.

Library Collections and the Crisis in Scholarly Communication

Since the publication of *University Libraries and Scholarly Communication* (Cummings et al. 1992), the scholarly communications system has become an increasingly visible conceptual framework within which the traditional practices of library collection development are being rethought. This is for two reasons, at least. First is the budgetary challenge to academic libraries that the report did so much to document and publicize. Increasing output of the scholarly publishing apparatus together with increasing unit costs in scholarly journals far exceed traditional budget allocations of universities to their research libraries. These increases have resulted in a well-publicized drop in the numbers of monographs and journals collected by research libraries (see, for example, Kyriillidou 2000) and, accordingly, diminished

access to this literature by researchers and students who depend on libraries. The effects of cost increases have been exacerbated, moreover, by sharp increases in the amount of published scholarly literature. Projecting from a hypothetical library that in 1980 could acquire all the world's published information, for example, Brian Hawkins factors together inflation in material costs, growth rates in publishing, and average rates of increase in research library budgets to conjecture that "available budgets in 2001 will only be able to purchase 2% of what they had twenty years before" and further that "collections will be archiving something of the order of one-tenth of 1% of the available information" (Hawkins 1998, 135). This sharp constriction of access to the body of peer-reviewed knowledge is variously known as the "crisis in scholarly communication" and the "crisis in scholarly publishing." (Two other dimensions of this crisis, in addition to the increase in unit costs and the increase in production, are the restrictions on permissible use imposed by many of the licenses that govern access to electronic books and journals, and the impermanence of digital information.)

A second reason for this new focus on scholarly communication—and a point perhaps more evident to us now, ten years after the publication of *University Libraries and Scholarly Communication*—is that the scholarly communications system itself is in the midst of a change that is unprecedented since its inception in the seventeenth century. If we define the scholarly communications system broadly to include the technological and institutional means by which theories, interpretations, and findings are submitted to the scrutiny of expert disciplinary communities and then critiqued, endorsed, disseminated, synthesized, and archived on behalf of a broad community of teachers and learners (novice and advanced, lay and professional), then the changes embedded in this commonplace observation are wide-ranging. Together with broader social changes in research and higher education, the application of the technologies of digital representation and networked communications to scholarly practice have resulted in broader participation in the scholarly communications process and wider access to scholarly information. E-mail has transformed direct scholar-to-scholar communication, vastly expanding opportunities for discourse and collaboration, particularly for scholars in smaller or more remote institutions. E-print servers have been established by working scholars and have expanded access to prepublished materials that previously circulated in mimeo and photocopy only within select circles. Electronic publication has begun to shorten the time between acceptance and distribution of refereed research. More significant, e-mail, datasets (in textual, visual, numeric, audio, and motion picture formats), preprints, and peer-reviewed work may now all be available in digital form, potentially accessible through a single

worldwide network. For the first time, all phases of the life cycle of scholarly work are potentially available to a global audience in an integrated database of knowledge. The potential worldwide social consequences of such expanded access to this body of peer-evaluated knowledge-claims can hardly be exaggerated, whether for facilitating scientific and technical advancement or for fostering cross-cultural understanding.¹

However, there is also significant risk in entrusting this knowledge to the fragile digital communications system. Although the technologies of digital representation and networked communication have often been seen as offering at least part of the solution for the cost crisis in scholarly publishing (cf., for example, Phelps 1998), such discussions typically give less attention to the equally momentous changes that are also transpiring within the business and organizational infrastructure of the information technology industry. Higher education and the scholarly communications system are connected more tightly than ever before to a network of commercial industries and services whose primary clientele is not necessarily the higher education community. These enterprises include the businesses that create information technology (hardware and software), commercial Internet services, overnight package delivery services, online bookstores, and document suppliers like ingenta (lately UnCover). Used in combination, such services have enabled university libraries to come closer to realizing a long-standing vision of a nationally or internationally distributed scholarly collection in which the responsibilities and the costs for acquiring, preserving, and delivering scholarly information can be shared among many repositories. These services have also significantly empowered scholars to access scholarly information on their own behalf without the mediation of libraries. However, the business environment within which these services exist—in particular, the information technology industry—is highly volatile, exposing the scholarly enterprise to greater risk of disruption than it is culturally prepared for. That is, the information products and services on which the scholarly communications system relies—including publishers, document suppliers, and hardware and software platforms—are increasingly under the control of commercial enterprises for which the opportunity and need to generate profit through competitive innovation increases the likelihood of business failure or product/service obsolescence. This economic volatility is not an accidental feature of the technology—or so I will propose (following Castells) later in this paper. Rather, it is an intrinsic characteristic of the socio-economic system Castells calls the "network society." The instability and volatility of this system, I suggest, need to be plainly acknowledged in our accounts of the crisis we face in the scholarly publishing system and its prospects for reform.

Market Economy, Academic Culture, and the Scholarly Communications Reform Movement

In his classic 1976 work *The Cultural Contradictions of Capitalism*, the sociologist Daniel Bell argues that to understand our postindustrial knowledge-based society we need to recognize the disjunctions that have emerged among its economy, its polity, and its culture. In his view, these three realms “are ruled by contrary axial principles: for the economy, efficiency; for the polity, equality; and for the culture, self-realization” (Bell 1976, xi–xii). “The techno-economic order,” he goes on to explain,

is concerned with the organization of production and the allocation of goods and services. . . . [Its] axial principle is *functional rationality*, and the regulative mode is *economizing*. Essentially, economizing means efficiency, least cost, greatest return, maximization, optimization, and similar measures of judgment about the employment and mix of resources (Bell 1976, 11).

By “culture,” on the other hand, Bell refers to “the realm of symbolic forms.” Bell argues that the endless demand within modernist culture for change and innovation has fallen out of step with the limitations imposed by economic realities: “changes in the economy and technology,” he writes, “are constrained by available resources and financial cost” while “changes in expressive symbols and forms . . . meet no resistance in the realm of culture itself” (Bell 1976, 34). Bell was specifically concerned with expressive symbols like paintings, poems, and musical composition. However, Bell’s analysis can be extended to the situation in academic research and scholarly publishing. That is, a similar disjunction can be said to exist between the academic culture within which symbolic representations of knowledge are produced—books and journal articles—and the economic structure that orders the dissemination of those representations.

The main features of the disjunction between academic culture and the economics of scholarly publishing have been suggested by Stevan Harnad (1995, 1999, 2001, *inter alia*), Corynne McSherry (2001), and several other writers.² Academic culture treats scholarly literature as if it were part of a gift economy: scholars donate their research and their writing to scholarly journals so that other scholars may read it, test it, teach it, and build upon it. Scholarly authors desire that their work be disseminated to as many interested readers—or as many intellectually qualified readers—as possible and that the barriers to readership therefore be kept as low as possible. However, scholarly writing and publishing are not—or are not simply—part of

a gift economy. Scholarly literature is also a product in a market economy. Universities and state and federal agencies pay for much of the infrastructure that supports this research, and the literature created by the collective research community is traditionally sold in published form—sometimes for modest amounts of money, sometimes for larger amounts—back to colleges and universities by commercial and non-commercial presses. All publishers must adjust prices and terms of distribution according to market conditions, and commercial publishers will generally adjust prices upward and restrict dissemination outside the circle of paying customers to the extent that law and the consumer market will permit.

Furthermore, the economic relationships between the libraries that are the primary purchasers of scholarly journals and the commercial publishers that have become the primary producers of those journals no longer constitute a well-functioning market. Instead, inelastic demand on the part of libraries has exacerbated the inflationary pressures created by horizontal integration within the publishing industry. Studies conducted by Mark McCabe (1998, 1999) indicate that the “sensitivity of library demand to price increases is very small by normal standards (a 1% increase in price results in a 0.3% decline in subscriptions). Given this inelastic demand, publishers have a strong incentive to increase prices faster than the growth rate of library budgets” (McCabe 1999). McCabe’s research also indicates that for journals sold by commercial publishers “prices are indeed positively related to firm portfolio size, and that [corporate] mergers result in significant price increases” (McCabe 1998). Finally, McCabe notes, “even after controlling for the effects of portfolio size and other variables, we still observed a substantial inflation residual” (McCabe 1999).

In response to this deterioration of market function, universities and university libraries have increasingly chosen to stimulate competition in pursuit of lower prices and less restrictive terms of access. The Scholarly Publishing and Academic Resource Coalition (SPARC), which was formed in June of 1998 by the Association of Research Libraries, is one of the most visible examples of this effort. SPARC is a worldwide alliance of research institutions, libraries, and other organizations that encourages competition in the scholarly communications market. Under the SPARC program, member libraries contribute to a capitalization fund that can be extended to not-for-profit scholarly publishers that choose to partner with SPARC. In addition, member libraries agree to select a certain number of SPARC-supported publications for subscription.

The central idea behind SPARC’s program for reform, according to the SPARC Enterprise Director Rick Johnson, “is that competitive market forces must be unleashed if the status quo [of high prices and restrictive access] is to be challenged. . . . Competition is the one overarching ideology

today that enjoys broad support among the disparate stakeholders in the scholarly communication process, including scientists in wide-ranging disciplines, librarians, administrators, and societies” (Johnson 1999). The logic behind the SPARC initiative, Johnson continues, is that “1) if authors have superior alternatives to existing high-priced journals, they will ultimately move to the outlet that better satisfies their need for both recognition and broad dissemination, and 2) if publishers have market support for bold (but inherently risky) new ventures, they are more likely to make the investment” (Johnson 1999).

SPARC and similar efforts are rational and potentially powerful responses on the part of the scholarly community to an economic crisis that threatens fundamental scholarly values. However, concern about the SPARC strategy may be raised on two related fronts. First, SPARC’s leverage is strongest on the consumer side of the market cycle: the libraries that agree to maintain a certain dollar value of annual subscriptions to SPARC products. This is the market component that is most motivated to create change. SPARC has so far had less impact on the producer side of the market, that is, the scholars who supply manuscripts and editorial services to scholarly journals. Without a deeper change in the motivational structures within scholarly culture, the ideology of competition will have limited effect.

At the same time, SPARC’s emphasis on digital technology to encode, store, and disseminate the products of scholarship to “improve the process of scholarly communication and reduce the costs of production and distribution” (2002) intertwines those scholarly products ever more tightly with software and hardware products and market cycles that will not be stabilized by the competitive pressures exerted by the SPARC community. That is, the cost of the scholarly products and the terms under which they may be used constitute only a part of the challenge we face as a scholarly community if enduring, wide availability of scholarship is our primary aim. The encoding schemes, file viewers, and network architecture that are necessary infrastructure for the use of the scholarly literature in digital formats will not be controlled by SPARC’s competitive strategies. Indeed, it might be expected that the targets of SPARC’s competitive products will be driven to innovate more intensively on the side of technology—forcing SPARC products to match those innovations in turn and thus increasing the economic and technological volatility that threatens the long-term availability of scholarship.

The scholarly community cannot insulate itself from the dynamic technological, economic, and social systems of which it is a part, and the reform efforts exemplified by the SPARC project are necessary and appropriate responses to the crisis in scholarly communication. Indeed, not to respond to market dysfunction and other changes in publishing and access would itself entail significant risk.

Nevertheless, the consequences of innovation and market intervention cannot all be predicted, especially in the tightly integrated realm of digital communication, and the associated risks cannot be avoided. However, these risks can be controlled and learned from only if the conditions from which they arise are kept in plain view.

Time, Space, and Libraries: Manuel Castells on the Information Technology Paradigm

Viewing the crisis in scholarly publishing from Daniel Bell’s perspective, we can say that it is not primarily technological or even economic, but instead that it is rooted in an academic or scholarly culture that relies on a market economy to distribute its products but that generally disavows the tools of that economy for setting the terms of distribution. Resolution of the crisis, it may be argued, will therefore require a *cultural* reorientation of the academy, cultivation of new scholarly and administrative practices focused on the management of the literature through attention to the consequences of choosing a particular venue for publication or providing peer review and other editorial services. Scholarly authors must recognize that if they choose to participate in a market economy, their actions will have market effects and, further, that as authors they can control some of those effects by selecting publishers according to their economic policies and practices. In similar fashion, libraries have begun to understand themselves as consumers that can exert market pressure on publishers by refusing to buy products whose cost—calculated in dollars or in the obligations or limitations imposed by contractual terms—exceeds use-value.

The cultural and economic realms whose disjunction Bell diagnosed are not entirely separate, of course, and the most visible initiatives in the scholarly communications reform movement may be interpreted as using economic tools to create exactly this kind of cultural change. More specifically, it might be argued that by intervening directly in the scholarly publishing market and creating new journals to compete with titles whose high-price/high-increase histories have created the budgetary crisis—and at the same time exploiting digital technologies that promise to enhance productivity and keep costs low—we are already presenting scholars with a set of choices that will help them recognize their impact on the economic infrastructure of the scholarly publishing system. Moreover, the SPARC-affiliated programs *Create Change* (Create Change 2000) and *Declaring Independence* (SPARC 2000) target the practice of working scholars more directly by providing them with checklists of responsible practices for scholarly publishing and asking them to evaluate the journals in which they publish against those guidelines.

As I have suggested, however, representational technologies are not culturally neutral: the material forms of information storage and transmission condition the cultural practices of the communities that use them. We should therefore expect that in adopting new technologies the traditional structures that have organized scholarly communication in the past will be subject to change. In his seminal three-volume study *The Information Age: Economy, Society and Culture*, the Spanish sociologist Manuel Castells proposes that digital technology has begun to “reshape, at accelerated pace, the material basis of society.”³ Economies throughout the world have become globally interdependent, introducing a new form of relationship between economy, state, and society, in a system of variable geometry” (Castells 2000, 1). In this new socioeconomic construction, according to Castells, the action of knowledge on knowledge—rather than the action of knowledge on raw material or on machines—has become the main source of economic productivity, and the symbolic representation of knowledge thus becomes central to social and cultural change. In this respect, the role of libraries as one of the key mediating institutions for the transmission and preservation of these symbolic representations places them at the center of the information society. However, this same action of knowledge on knowledge made possible by digital technology creates unprecedented flux in the technical, economic, and social infrastructure through which libraries perform this role, thus challenging their ability to fulfill their traditional mission and motivating changes in the way the work of libraries is organized.

In this section, I will outline Castells’s theory of the relationship between what he calls the “information technology paradigm” and emerging features of early twenty-first-century culture. Castells argues that “the cumulative feedback loop between innovation and the uses of innovation” (Castells 2000, 31) made possible by networked digital technology leads to increasingly rapid product innovation and increasing volatility in business cycles. And because this new economy is based on technologies that represent knowledge and information, Castells argues, we should expect social change as well—the emergence of new forms of social organization and cultural production. Among these changes is a change in the way time organizes the relationships of work, family, and other social groups. This is an intensification, I will argue, of a basic feature of modernity that Anthony Giddens (1990) calls “disembedding,” in which social interaction is dispersed across time and space rather than being localized in time and place. In the section that follows, I will connect Castells’s work more closely to Giddens’s broader theory of the dynamic of modern social systems and the defining place of risk in those societies and will suggest that the technological and economic structures that define the “information society”—

and the knowledge objects bound into those structures—are inherently at risk of dissolution. I will then draw some implications for the practice of librarians and the scholarly communications reform movement.

In the new “informational” mode of development, according to Castells, the source of economic productivity has become the application of technology to knowledge generation, information processing, and symbolic communication (Castells 2000, 17). (A mode of development, in Castells’s terms, is the set of “technological arrangements through which labor works on matter to generate the product, ultimately determining the level and quality of surplus” [16]). By contrast, in an agrarian mode of development, “the source of increasing surplus results from quantitative increases of labor and natural resources (particularly land) in the production process, as well as from the natural endowment of these resources,” whereas in an industrial mode of development, “the main source of productivity lies in the introduction of new energy sources, and in the ability to decentralize the use of energy throughout the production and circulation processes” (16–17).

What is new in the informational mode of development, Castells explains, is not the kind of activities in which humankind is engaged, but rather “its technological ability to use as a direct productive force . . . [its] capacity to process symbols” (100). Thus for Castells what is specific to the informational mode of development “is the *action of knowledge upon knowledge* itself as the main source of productivity. . . . Information processing is focused on improving the technology of information processing as a source of productivity” (17, emphasis added). Modes of development shape social behavior, including symbolic communication; and because informationalism is based on technologies that represent knowledge and information, Castells argues, “there is an especially close linkage between culture and productive forces, between spirit and matter, in the informational mode of development. It follows that we should expect the emergence of historically new forms of social interaction, social control, and social change” (18).

When we talk about information technologies, we too often think only of digital technology and overlook the complex analog technologies embodied in writing, drawing, printing, and other forms of representation and storage of information and knowledge. Castells argues, however, that the tightly integrated technologies of digital representation and networked (packet-switching) communication make a *specific* difference to the informational mode of development and thereby influence culture and cultural practices.⁴

Castells’s analysis thus covers both the economic and the cultural consequences of the informational mode of development and may help us to see how the role of libraries in the preservation and transmission of knowledge may be changed by these new cycles of technological innovation.

On the one hand, Castells argues, information technology is itself now the engine of economic growth: it enables the iterative application of knowledge to improve the technology. In turn, information technology enables the globalization and concentration of capital by creating networks of information that “converge toward a meta-network of capital that integrates capitalist interests at the global level and across sectors and realms of activity” (506). High-technology firms depend on those highly concentrated financial resources to sustain their endless drive toward innovation, productivity, and competitiveness; and capital, “acting directly through financial institutions or indirectly through the dynamics of stock exchange markets, [thereby] conditions the fate of high-technology industries” (503–4). Manipulation of these financial markets—instantaneous shifts of large sums of capital in response to equally instantaneous communication of changing political and economic circumstances, practices made possible by electronic networks—in turn creates new forms of devastating economic crises, leading to the “wrecking of companies, and of their jobs, regardless of performance, because of sudden, unforeseen changes in the financial environment in which they operate” (466).

For Castells, therefore, the information technology industry and the larger society of which it is a part are inevitably and necessarily unstable and subject to crisis. “Any attempt at crystallizing the position in the network as a cultural node in a particular time and space sentences the network to obsolescence, since it becomes too rigid for the variable geometry required by informationalism” (215). The root of this condition of instability, according to Castells, is the relationship of network technology to time:

[D]uring the 1990s the convergence of global deregulation of finance and the availability of new information technologies and new management techniques transformed the nature of capital markets. For the first time in history, a unified global capital market, *working in real time*, has emerged. The explanation, and the real issue, of the phenomenal volume of trans-border financial flows . . . lies in the *speed* of the transactions. The same capital is shuttled back and forth between economies in a matter of hours, minutes, and sometimes seconds (465).

At the same time, Castells claims, the processes of social transformation within the network society go beyond the sphere of social and technical relationships of production and affect culture and power as well. In the network society, “[c]ultural expressions are abstracted from history and geography, and become predominantly mediated by electronic communication networks” (507). One of the key

categories of cultural expression, according to many social theorists, is the way time organizes work and other social relationships and processes.⁵ Lewis Mumford, for example, argues that the “clock, not the steam-engine, is the key-machine of the modern industrial age” (Mumford 1934, 14) for making possible the rational organization and coordination of social and industrial enterprise.

A society structured around technologies of digital networking, Castells claims, is characterized by the breaking down of both the traditional biological and social rhythms associated with the notion of a life cycle and the clock time of industrial society (476). He calls this a condition of “timeless time”—a condition that “occurs when the characteristics of . . . the informational paradigm and the network society, induce systemic perturbations in the sequential order of phenomena. . . . Elimination of sequencing creates undifferentiated time” (494). Castells argues that

this is happening now not only because capitalism strives to free itself from all constraints, since this has been the tendency of the capitalist system all along, without being able fully to realize it. Nor is it sufficient to refer to the cultural and social revolts against clock time, since they have characterized the history of the past century without actually reversing its domination, indeed furthering its logic by including the clock time distribution of life in the social contract. *Capital's freedom from time and culture's escape from the clock are decisively facilitated by new information technologies, and embedded in the structure of the network society* (464; emphasis supplied).

Similarly, Castells argues, spatially localized *places* are giving way to what he calls the “space of flows.” “From the point of view of social theory,” Castells explains, “space is the material support of time-sharing social practices” (441). In other words, space brings together social practices that are simultaneous in time. Traditionally, this “bringing together” was accomplished by physical contiguity or proximity. In the case of libraries, for example, it has meant that for much of their history one of the motivations for building large comprehensive print collections has been to control the inconvenience caused by spatial dispersion of information-bearing documents; effective access to information required spatial proximity to the documents in which it was embodied.

In the informational mode of development, physical proximity is being replaced by other kinds of material supports for simultaneous social practices—circuits and networks of electronic exchanges and the nodes and hubs that organize these networks (442–43)—and society is increasingly constructed around what Castells calls “flows,” pur-

poseful and repetitive sequences of exchange and interaction between physically separated actors in the economic, political, and symbolic structures of society. Capital may be exchanged along these flows as may information, technology, organizational interaction, images, etc. Castells proposes that a “space of flows” is the “new spatial form characteristic of social practices that dominate and shape the network society. . . . *The space of flows is the material organization of time-sharing social practices that work through flows*” (442). We can accordingly think of libraries not as individual places or structures but as nodes within a space of information flow, a space in which simultaneous access to information objects is not necessarily accomplished through physical proximity (of the objects or the user).

The upshot, Castells argues, is that networked digital technology is beginning to reconfigure the most basic structures of society and culture. It is important to note that “timeless time” and the “space of flows” are not simply psychological categories, not merely the ways in which some members of the informational society have come to experience their world. Rather, these new organizing principles have material consequences. As I shall argue in the next section, they help to create a condition of standing or intrinsic risk not characteristic of earlier forms of modernity.

From Technological Determinism to Risk Culture: Implications for Library Collection Development

Castells has been criticized for the technological determinism that appears to inform this account of the relation between modes of production (capitalism) and modes of development (informationalism) (cf., for example, van Dijk, n.d.). That is, it may appear from Castells’s account that the social and symbolic aspects of our lives are shaped exclusively by the conditions of high technology and late capitalism, that technology is a juggernaut out of our control carrying society to its inevitable destiny (glorious or debased, as the case may be).⁶ In this section, I want to propose an alternative interpretation by which Castells’s work is read instead as describing a condition of risk that is inherent in the informational mode of development. Under this reading, the deterministic tendency in Castells’s account is counterbalanced by Anthony Giddens’s concept of reflexivity in the risk society, and the conditions that Castells identifies can be addressed through political or social action.

Unlike Castells and other theorists, Giddens does not treat “informationalism” or “network society” as forms of society that have radically broken with Western modernity. Rather, he considers modernity to have entered a period of extreme intensification, and the “timeless time” and “space

of flows” in Castells’s account of network society can thus be understood as an extreme condition of what Giddens calls “disembedding,” “the ‘lifting out’ of social relations from local contexts of interaction and their restructuring across indefinite spans of space-time” (Giddens 1990, 21). Disembedding, in Giddens’s view, is one of the hallmarks of modern culture:

The dynamism of modernity derives from the *separation of time and space* and their recombination in forms which permit the precise time-space ‘zoning’ of social life; the *disembedding* of social systems (a phenomenon which connects closely with the factors involved in time-space separation); and the *reflexive ordering and reordering* of social relations in the light of continual inputs of knowledge affecting the actions of individuals and groups (16–17).

Giddens contrasts the disembeddedness characteristic of modernity with the tighter integration of time and place characteristic of more traditional societies. In premodern societies, he says, “space and place largely coincide, since the spatial dimensions of social life are, for most of the population, and in most respects, dominated by ‘presence’—by localised activities. The advent of modernity increasingly tears space away from place by fostering relations between ‘absent’ others, locationally distant from any given situation of face-to-face interaction” (18).

The various technologies and social practices that create spatial and temporal disembedding—including long-distance communication technologies—have allowed modern bureaucratic organizations such as universities (and also states and corporations) to coordinate the activities of large numbers of people across large regions of space and long periods of time, thus connecting “the local and the global in ways which would have been unthinkable in more traditional societies and in so doing routinely affect the lives of many millions of people” (20). As a consequence, the potential reach of unintended by-products or technical failures of these technologies and practices is greatly magnified in the numbers of people who may be affected and the size of the regions across which the consequences may spread. For example, whereas the tendency of deadly viruses and other biological agents to kill their hosts can be a significant limitation on their ability to spread widely, the high-speed transportation networks that have helped to reduce the effects of spatial dispersion (by reducing the amount of time previously required to move people and products across great distances) are now more likely to disperse those viruses to major population centers.

Giddens thus characterizes the modern condition as one of intrinsic risk, risk that is created by the social practices and

technologies of modern life as contrasted with the kinds of dangers presented by life in the premodern world. Moreover, he says, modernity is also characterized by widespread *awareness* not only of the risks we face but also of the limitations of scientific and technical expertise in controlling or resolving those risks. This awareness is an aspect of what Giddens calls the “reflexivity” of modern life (124 ff.). “There is a fundamental sense,” Giddens explains, “in which reflexivity is a defining characteristic of all human action” (Giddens 1990, 36). “All human beings routinely ‘keep in touch’ with the grounds of what they do as an integral part of doing it” (Giddens 1990, 36). With the advent of modernity, however, reflexivity takes on a different character. “It is introduced into the very basis of system reproduction, such that thought and action are constantly refracted back upon one another. . . . The reflexivity of modern social life consists in the fact that social practices are constantly examined and reformed in the light of incoming information about those very practices, thus *constitutively* altering their character” (38; emphasis supplied).

Not only are social systems constantly evaluated by various groups of experts and laypeople, the resulting reflexive knowledge is itself reflexively used to modify those systems and thus the nature and dynamics of modern social systems. Indeed, in a network society as described by Castells, we should expect to see an intensification of that loop of information-modification-information.

Giddens and other theorists of the so-called “risk society” (cf. Beck 1992) have generally focused on threats such as nuclear annihilation, environmental collapse, and worldwide contagion from genetically modified organisms, and have proposed political and social responses appropriate to these dark realities. Castells’s account of the fundamental structural role of information in the network society suggests that the fragility of digital networks may pose a risk with similar reach. If the “action of knowledge upon knowledge” is the fundamental source of productivity in the information economy, then threats to the continuing accessibility of the body of validated knowledge are equally threats to the sustainability of this economy. Moreover, we may hope that this body of knowledge also contains the tools necessary for moderating some of the risks that advanced science and technology have helped to create; the loss of that literature would therefore be all the more tragic.

One of the mechanisms that create disembedding in modern societies is the “expert system,” by which Giddens means “systems of technical accomplishment or professional expertise that organise large areas of the material and social environments in which we live today” (27). Expert systems “remove social relations from the immediacies of context . . . by providing ‘guarantees’ of expectations across distanced time-space” (28). Academic research libraries can be understood as a disembedding expert system, as may

many other elements of the scholarly communications system. In an academic library, the labor of economic consumption of scholarly literature is divided from the labor of production. Professional librarians take responsibility for the business processes by which the literature that is created by and for the use of scholars is bought and paid for, and also take responsibility for some aspects of the organization and long-term storage of that literature. The principles and processes by which we accomplish this work define our professional expertise.

However, in other ways universities and university libraries still embody to an unusual degree the more traditional “premodern” integration of time and space and of space and place, while scholarly publishing combines aspects of both a traditional “face-to-face” culture and a more abstract market-driven system. As we noted in the previous section, for example, one of the motivations for building large, comprehensive print collections has been to reduce the inconvenience caused by spatial dispersion of information-bearing documents; effective access to information has required spatial proximity to the documents in which it is embodied, and many of the expectations and practices of academic workers are structured around the local print repository. The crisis in scholarly communications, understood as a cultural crisis, may therefore be traced to the tension between the traditional aspects of scholarly practice—research, authorship, and peer review—and the more modern or abstract systems that result in the pricing and marketing of commercial journals and other scholarly publications.

Until recently (and the establishment of the annual ritual of serial cancellation at most universities) most scholars have not had to directly confront the market behavior of the journals that, as authors and readers, they support. Similarly, most scholars have not had to confront the consequences for themselves and their community of the restrictions imposed by many publishers on the use of scholarly literature in digital networks. At the least, this division of responsibility between librarians and scholars has contributed to the perception on many campuses that inflation, underfunding, and licensing restrictions are a library problem.

If librarians wish to contribute to the resolution of the scholarly publishing crisis, therefore, we may need to return more responsibility for the functions of that system to the scholars who create and consume its products. The *Create Change* program sponsored by SPARC, the normalized pricing studies undertaken by Cornell (Cornell University Faculty Task Force 1998) and the University of Wisconsin (Soete and Salaba 1999), and other efforts on local campuses to inform faculty about the consequences of publisher practice (cf., for example, Fyffe and Kobulnicky 1999) represent some of the steps necessary for creating a critical self-understanding on the part of scholarly authors and readers—for increasing, as Giddens would call it, the

“reflexivity” of the academy, thereby helping to resolve the contradiction between the traditionalist and modernist aspects of the scholarly communications system. However, having taken professional responsibility for much of the apparatus of the organization of scholarly literature and for some aspects of its distribution, professional librarianship has instead helped to block reflexive feedback and the changes in the respective roles of authors and librarians that this might entail.

There are limits, of course, to the degree to which reflexive knowledge—on the part of librarians or that of scholarly authors and readers—can control the changes introduced by new technologies. Resolution of the cost crisis in scholarly publishing will not eliminate the susceptibility of digital systems to technical failure nor reduce the interest of commercial publishers in using mechanisms other than cost to restrict the availability of the intellectual property under their control (restrictive licenses, for example). The new landscape of risk is therefore one with which librarians and scholars alike must become more familiar. Even traditional print-based libraries did not have absolute control over their services; outside forces were always capable of disrupting the delivery schedules of books and journals, and individual copies of printed books and journal issues could always be lost or vandalized. However, as digital networking, fax transmission, and rapid package delivery begin to offer an alternative to local collection development for meeting access needs, the increasing dependence of scholarly communication on these businesses, systems, and technologies weakens the scholarly community’s control over the scholarly communications system and leaves it more vulnerable to highly disruptive change. The denser and further flung the network from which the library delivers its services—particularly the scholarly texts and other information for which libraries are the traditional repository—the greater the risk of service disruption. Despite the wishes and expectations of some library users, the reliability of access strategies like document delivery and remotely hosted digital files cannot be guaranteed by the local institution. Indeed, Castells’s observations on the relationship between competition in the information industry and the large-scale flows of capital reinforce the common-sense expectation that the volatility of these markets and services will only intensify in the coming years.

As with the costs of scholarly information, however, librarians have tended to mask the volatility of information services in an effort to create “seamless” or “transparent” systems. Instead, I would suggest, it is vital that the faculty and students for whom library services are designed as well as the administrators responsible for funding those services be helped to understand the increased risk and volatility inherent in the transformed scholarly communications network. One of the means by which this awareness can be increased is the collection development policy. In a 1986

paper, Ross Atkinson analyzes the functions of collection policies into referential, generative, and rhetorical functions. The referential function is primary, he says; it “provides a description of the collection’s current state, development, and desired direction” (141). The generative function, in which the policy guides the selector in transforming the collection from its current to its desired condition, and its rhetorical function, in which it provides an argument “that there is a systematic collection plan in effect, and that such a plan is worth pursuing” (141), follow, he says, from the policy’s referential function.

The new conditions of the access library and the crisis in scholarly communications of which these conditions are an aspect argue for a re-ordering of these priorities with greater prominence given to the rhetorical function. The traditional collection development policy needs to be reconceived as a strategically oriented access-development plan guided by the transformations under way in the scholarly communications system. Such a plan should articulate, for each disciplinary program, the roles that local collections, remotely hosted digital files, and document delivery services will play in providing information. Such a plan should also highlight the sources of risk to the short-term and long-term availability of information under these models, as conditioned by rising costs, access restrictions imposed by owners of the intellectual property, volatility among the key publishers, etc. In the unstable state in which that system currently finds itself, the rhetorical function of the access plan therefore takes on greater importance. The stability, rationality, and predictability of information markets on which a “systematic collection plan” would be founded are not ours to claim, and it is vital to our credibility that we articulate the limits of our control over information services.

The stakes for libraries, if they are to remain an integral part of the scholarly communications system, are high. As Giddens points out, “Widespread lay knowledge of modern risk environments leads to awareness of the limits of expertise and forms one of the ‘public relations’ problems that has to be faced by those who seek to sustain lay trust in expert systems” (1990, 130). However, while the existence of risk poses a threat to the credibility of experts, it is worse for an expert community to be discovered to have concealed risk or to have ignored it altogether:

The faith that supports trust in expert systems involves a blocking off of ignorance of the lay person when faced with the claims of expertise; but the realization of the areas of ignorance which confront the experts themselves, as individual practitioners and in terms of overall fields of knowledge, may weaken or undermine that faith on the part of lay individuals. Experts often take risks “on behalf” of lay clients while concealing, or fudging over, the true

nature of those risks or even the fact that there are risks at all. More damaging than the lay discovery of this kind of concealment is the circumstance where the full extent of a particular set of dangers and the risks associated with them is not realised by the experts (Giddens 1990, 130–31).

Conclusion

One of the functions of social theory is to help bring about new ways of viewing familiar phenomena. When an abstract story is created about the details and complexities of everyday social practices, new connections may be revealed between areas not previously seen as connected. In this paper, I have attempted to draw connections between the reform efforts currently under way in the areas of scholarly publishing and scholarly communication, on the one hand, and theories of the emerging shape of risk in societies structured by information technology and networked communication, on the other. I have suggested that risk of loss of scholarly knowledge be understood as an intrinsic feature of digital information technology, not as an accidental limitation that will eventually be overcome; and that some of the efforts currently under way to reform the scholarly communications system may increase that risk by increasing the instability of the scholarly publishing market. I have argued, moreover, not that we should seek to avoid risk, but instead (the risk being unavoidable) that these risks need to be made clear to the scholarly communities served by librarians, and that greater responsibility for the choices presented by evolving information services should be returned to the scholarly community that creates and uses the scholarly literature. There is significant risk for librarians, I concluded, in accepting risk “on behalf of” the community we serve, unless those risks are clearly explained and articulated.

Making such adjustments to the division of scholarly labor created by the modern bureaucratic university will not come easily or quickly. There are clear benefits to this division that we should wish to preserve. However, we should also expect that the changes under way in the emergence of the “network society” or “information society” will include the traditional roles and relationships of scholarly authors and academic librarians.

Notes

1. Compare Lewis Mumford on the historic impact of the experimental method in science:

[T]he most important invention of all had no direct industrial connection whatever: namely, the invention of the experimental method in science. This was without

doubt the greatest achievement of the eotechnic phase [Mumford’s term for the handicraft technology prior to the industrial revolution]: its full effect upon technics did not begin to be felt until the middle of the nineteenth century. The experimental method . . . owed a great debt to the transformation of technics: for the relative impersonality of the new instruments and machines, particularly the automata, must have helped to build up the belief in an equally impersonal world of irreducible and brute facts, operating as independently as clockwork and removed from the wishes of the observer: the reorganization of experience in terms of mechanical causality and the development of cooperative, controlled, repeatable, verifiable experiments, utilizing just such segments of reality as lent themselves to this method. . . . None of the inventions that followed the development of the scientific method were so important in remolding the thought and activity of mankind as those that made experimental science possible (Mumford 1934, 132–33).

2. Harnad argues that scholarly authorship is properly part of a gift economy and must therefore be carefully distinguished from commercial authorship that is part of a market economy. Copyright protections are appropriate for the latter, he says, but not for the former. Harnad therefore proposes that scholarly publishing be reorganized around a system of freely available open-access archives of scholarly literature. McSherry, by contrast, emphasizes the close ties between university funding and the commercial marketplace. Even so, they are both skeptical—at least with respect to academic work—that “the natural compensation for creative work is property ownership” (McSherry 2001, 26). For further discussion of the relationship between universities and the commercial economy, see Slaughter and Leslie (1997).
3. For a useful critical overview of Castells’s trilogy, see Stalder (1998).
4. Castells focuses specifically on digital (binary) schema for representing knowledge objects and on packet-switching as a communications protocol. However, most features of his analysis would stand if digital representation were replaced by some other system and if (when) other communications protocols emerge. What is fundamental to Castells’s analysis is any representational schema that can integrate multiple sensory modes into a common high-speed communications channel.
5. I have here reversed Castells’s order of exposition. In contrast to most social theorists, Castells considers space to be a more fundamental organizing principle than time. For the purposes of this paper, however, the point is not crucial.
6. For an overview of historical determinism, see Smith and Marx (1994). Feenberg (1999) offers a useful antideterminist theory of contemporary technology and society.

Works Cited

- Atkinson, Ross. 1986. The language of the levels: Reflections on the communication of collection development policy. *College and Research Libraries* 47, no. 2: 140–49.

- Atkinson, Ross. 1996. Library functions, scholarly communication, and the foundation of the digital library: Laying claim to the control zone. *The Library Quarterly* 66, no. 3: 239–65.
- Atkinson, Ross. 2000. A rationale for the redesign of scholarly information exchange. *Library Resources and Technical Services* 44, no. 2: 59–69.
- Beck, Ulrich. 1992. *Risk society: Towards a new modernity*, trans. Mark Ritter. London: Sage Publications.
- Bell, Daniel. 1976. *The cultural contradictions of capitalism*. New York: Basic Books.
- Castells, Manuel. 2000. The rise of the network society. In *The Information Age: Economy, society and culture*, vol. 1, 2d ed. Oxford: Blackwell.
- Cornell University Faculty Task Force. 1998. *Journal Price Study of Core Agricultural and Biological Journals*. Accessed Aug. 9, 2001, <http://jan.mannlib.cornell.edu/jps/jps.htm>.
- Create Change. 2000. Accessed Aug. 9, 2001, www.createchange.org.
- Cummings, Anthony M., et al. 1992. *University libraries and scholarly communication: A study prepared for the Andrew W. Mellon foundation*. Washington: Association of Research Libraries.
- Eisenstein, Elizabeth. 1979. *The printing press as an agent of change: Communications and cultural transformations in early modern Europe*. Cambridge: Cambridge Univ. Pr.
- Eisenstein, Elizabeth. 1983. *The printing revolution in early modern Europe*. Cambridge: Cambridge Univ. Pr.
- Feenberg, Andrew. 1999. *Questioning Technology*. London: Routledge.
- Fyffe, Richard C., and Paul J. Kobulnicky. 1999. Negotiating the soul of the library: Change management in information access and local collection development. *Journal of Library Administration* 28, no. 4: 17–35.
- Giddens, Anthony. 1990. *The consequences of modernity*. Stanford: Stanford Univ. Pr.
- Harnad, Stevan. 1995. The post-Gutenberg galaxy: How to get there from here. *The Information Society* 11, no. 4: 285–92. Accessed Aug. 9, 2001, <http://cogsci.soton.ac.uk/~harnad/THES.thes.html>.
- Harnad, Stevan. 1999. Free at last: The future of peer-reviewed journals. *D-Lib Magazine* 5, no. 12. Accessed Feb. 8, 2002, www.dlib.org/dlib/december99/12harnad.html.
- Harnad, Stevan. 2001. For whom the gate tolls? How and why to free the refereed research literature online through author/institution self-archiving, now. Accessed Aug. 9, 2001, <http://cogsci.soton.ac.uk/~harnad/Tp/resolution.htm>.
- Hawkins, Brian L. 1998. The unsustainability of the traditional library and the threat to higher education. In *The mirage of continuity: Reconfiguring academic information resources for the 21st century*, Brian L. Hawkins and Patricia Battin, eds. Washington, D.C.: Council on Library and Information Resources and the American Association of Universities.
- Innis, Harold. 1972 (1950). *Empire and communications*. Revised by Mary Q. Innis. Toronto: Univ. of Toronto Pr.
- Johnson, Richard. 1999. Competition: A unifying ideology for change in scholarly communications. Accessed Aug. 9, 2001, www.arl.org/sparc/core/index.asp?page=a2.
- Kyrillidou, Martha. 2000. Journal costs: Current trends and future scenarios for 2020. *ARL: A Bimonthly Report on Research Library Issues and Actions*, no. 210 (June). Accessed Aug. 9, 2001, www.arl.org/newsltr/210/costs.html.
- MacLean, Margaret, and Ken H. Davis. 1998. *Time and bits: Managing digital continuity*. Los Angeles: J. Paul Getty Trust.
- McCabe, Mark. 1998. The impact of publisher mergers on journal prices: A preliminary report. *ARL: A Bimonthly Newsletter of Research Library Issues and Actions*, no. 200 (Oct.). Accessed Aug. 9, 2001, www.arl.org/newsltr/200/mccabe.html.
- McCabe, Mark. 1999. The impact of publisher mergers on journal prices: An update. *ARL: A Bimonthly Newsletter of Research Library Issues and Actions*, no. 207 (Dec.). Accessed Aug. 9, 2001, www.arl.org/newsltr/207/jmlprices.html.
- McLuhan, Marshall. 1962. *The Gutenberg galaxy: The making of typographical man*. Toronto: Univ. of Toronto Pr.
- McSherry, Corynne. 2001. *Who owns academic work? Battling for control of intellectual property*. Cambridge: Harvard Univ. Pr.
- Mumford, Lewis. 1934. *Technics and civilization*. New York: Harcourt, Brace, and Co.
- Phelps, Charles. 1998. Achieving maximal value from digital technologies in scholarly communication. *ARL Proceedings* 133. Accessed Aug. 9, 2001, www.arl.org/proceedings/133/phelps.html.
- Slaughter, Sheila, and Larry L. Leslie. 1997. *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: The Johns Hopkins Univ. Pr.
- Smith, Merrit Roe, and Leo Marx, eds. 1994. *Does technology drive history? The dilemma of technological determinism*. Cambridge: MIT Pr.
- Soete, George, and Athena Salaba. 1999. *Measuring the cost-effectiveness of journals: Ten years after Barschall*. Accessed Aug. 9, 2001, www.library.wisc.edu/projects/glsdo/cost.html.
- SPARC. 2000. *Declaring independence: A guide to creating community-controlled science journals*. Washington, D.C.: Scholarly Publishing and Academic Resources Coalition. Accessed Aug. 9, 2001, www.arl.org/sparc/di.
- SPARC. 2002. FAQ. Accessed July 26, 2002, www.arl.org/sparc/core/index.asp?page=e0.
- Stalder, Felix. 1998. The network paradigm: Social formations in the age of information. *The Information Society* 14, no. 4: 301–8.
- van Dijk, Jan A. G. M. n.d. The one-dimensional network society of Manuel Castells. *Chronicle World*. Accessed May 26, 2001, www.thechronicle.demon.co.uk/archive/castells.htm.
- Waters, Donald J. 1999. Steps toward a system of digital preservation: Some technological, political, and economic considerations. In *The mirage of continuity: Reconfiguring academic information resources for the 21st century*, Brian L. Hawkins and Patricia Battin, eds. Washington, D.C.: Council on Library and Information Resources and Association of American Universities.

Inventory at Brooklyn College, 1998–1999

An Original Method

Judith W. Wild

This article discusses the development of an inventory project at Brooklyn College that entailed examining the collection and comparing it to the corresponding records in the online catalog. The procedure became necessary in large part due to problems resulting from the migration to a new, integrated cataloging system in 1987. We needed to deal with (a) books in the catalog that were not on the shelves, (b) books on the shelves that were not in the catalog, and (c) books that lacked circulation information (item records). We used the circulation module of our integrated system to discharge every book, thereby changing its record. An unchanged record indicated a missing book. Missing books were then removed from the catalog. Books on the shelves with no bibliographic record were redeemed and entered into the catalog. Item records were created for those books that needed them. Other errors were also identified and corrected during this time.

Introduction and Rationale

In 1998, the Brooklyn College Library began an inventory of the main circulating monograph collection known in our NOTIS catalog as “Brooklyn Stacks.” The chief impetus to embark on such an undertaking was the expansion and renovation of the library, slated to begin in August of 1999. For at least one year, we would work in temporary quarters with the library’s holdings in closed stacks. Because the books could not be browsed, it became crucial that the catalog be accurate. A catalog that failed to match the collection would adversely affect the paging service, as aides would be sent to fetch nonexistent books. Such a situation would be frustrating for our patrons, who would need to research and resubmit their requests, and expensive for the administration funding this service. Also, we knew that after moving the books first to temporary quarters, it would be necessary to send them back, which made it important for us to know exactly what was on the shelves. We had to ascertain which titles, if any, had been misshelved or lost during the course of the moves. We selected only the Brooklyn Stacks collection of about 500,000 books because it was the largest circulating collection.

As far back as 1995, when the library underwent an outside evaluation, the chief librarian made it known that she felt a collection inventory was a high priority. In my capacity as head of technical services, I was charged with looking into the matter and was not entirely surprised to learn that, like ours, many libraries badly needed an inventory but lacked both the human and financial resources to undertake such a commitment. When it became clear that the library construction project would indeed move forward, we received approval from the then vice president of finance and administration to pursue the inventory. The question

Judith W. Wild
(jwild@brooklyn.cuny.edu) is Associate Professor/Head of Technical Services, Brooklyn College Library, Brooklyn, N.Y.

Manuscript received January 4, 2002; accepted February 25, 2002.

came down to how it could be accomplished. Due to problems in our catalog, we turned to a unique method that used the circulation module to change the records' status for the books on the shelves.

To understand the daunting work ahead of us, a brief history is in order. In 1975, the last inventory of the main circulating collection had been attempted but not completed, so we knew we faced a formidable task (Yu 1997). Furthermore, from circulation staff statistics on "searched but not found books," it was clear that many works could not be located. Such statistics are compiled manually when a reader turns to the circulation desk for assistance after unsuccessfully looking through the stacks for a book listed in the catalog as available; if the circulation staff, after searching the surrounding area, cannot find the book either, it is added to the count. To make matters worse, many titles on the shelves were not in the NOTIS catalog. Although these titles had been in the card catalog, they had never been converted in the Microcon (retrospective conversion of books in Library of Congress Classification) project.

Moreover, although a great deal of effort had been made to correct the problem, most of our bibliographic records did not have linked item records. An item record contains the book's bar code so that it can be checked out. Also included is the circulation history, indicating whether the book is on the shelf or checked out, as well as the last date of check out and return. In the NOTIS system, a *linked* item record is an item record that is attached to the bibliographic record and shares its bibliographic data. An *unlinked* item record is an item record that is not attached to the bibliographic record, has limited bibliographic information, and is only accessible in staff mode through its NOTIS number or bar code.

If, at the point that the book is being checked out, there is no item record, obviously, one must be created. Circulation staff do not always have the knowledge to add the item record to the correct copy. Thus, the process of creating an item record can produce both long lines at the circulation desk and mistakes. While unlinked item records are not accessible to the reader, they can be used by circulation staff to check the book in and out. In our situation, these books upon return were sent to cataloging for linking. We had many unlinked item records that had been created from the data in our CLSI database, the automated circulation system we used from 1982 to 1987. Linking the unlinked item records became the cataloging unit's interminable ordeal. More will be said about these records later.

Background

Brooklyn College is a member of the City University of New York (CUNY), a consortium of nineteen institutions.

The college, founded in 1930, had its first home in rented quarters in northern Brooklyn, locally known as "downtown" Brooklyn. In 1937, the college moved to the Midwood district, its present location in central Brooklyn (Brooklyn College Library 2001). In 1971 a satellite campus with its own library was set up in downtown Brooklyn; when it closed in 1976, the collections and records of the two centers had to be merged, a process that took many years to complete.

In 1974, the library began to use OCLC for cataloging but did not change its procedures for adding copies. They were typed directly onto the shelflist, not added to the OCLC record. Hence, our OCLC archive tape did not reflect copy information (Iskenderian 1997). This procedural decision may be difficult to understand at the present time, but at that time OCLC was in its infancy and we were still experimenting with it. In NOTIS, the default is one copy. Consequently, the absence of copy information meant that when our OCLC records were loaded into the NOTIS catalog, they all appeared as single copies.

The circulation function was automated in 1982 through our acquisition of the CLSI system. In December of 1987, along with John Jay and Baruch Colleges, Brooklyn College became one of the first libraries in the City University of New York to have an online catalog. This catalog, which we call CUNY+, is a NOTIS catalog and is still being used at the time of this writing. Plans are underway to switch to Ex Libris' Aleph system during the fall of 2002.

As a result of all these changes, the catalog had become compromised. We used OCLC's Microcon process to convert our pre-1974 records for Library of Congress classified books to machine-readable form. Thousands of these records failed to be converted and hence turned up on an exceptions list. This list consisted both of titles for which there were no corresponding records in OCLC and titles that had been keyed in incorrectly on the Microcon grid. Budgeting constraints necessitated my predecessor's decision not to add any of these exceptions to the online catalog until such time as an inventory would be performed, the reason being that he did not want us to spend time adding records for which there might not be books. For economic reasons, an inventory was not performed, so these unconverted LC records went untouched while we processed new books and continued to reclass even older books that were still in Dewey Decimal Classification. These neglected items, which are referred to as the "red dot" books, will be discussed later in this article.

Circulation Information

In preparation for NOTIS, an attempt was made to capture all bar code and circulation information created from 1982

to 1987 through CLSI residing in the system. These became the infamous unlinked item records. Unfortunately, the number of Brooklyn College records exceeded the capacity of the loader of the system to which we were migrating, and many were excluded (Bowdoin 1997).

From 1987 to 1998, the cataloging unit linked many unlinked item records from reports generated for us by the CUNY Office of Library Systems. In spite of these efforts, just before the inventory was performed, when I requested a *linked* item report from the CUNY office on the Brooklyn Stacks collection, I discovered that only 212,000 item records had been linked to bibliographic records. Given how many titles we suspected were missing and a rough count of what remained on the shelves (500,000), it was evident that far less than half of the collection had linked item records; more than likely, the report represented both missing books as well as those on the shelves. A catalog record without a linked item tells the reader nothing about the book's whereabouts; thus, the absence of an item record appropriately generates the message "check shelf." Whereas in an open stack environment this is not important, in a closed stack environment it would be unacceptable to ask readers to fill out paging slips if they could not ascertain from the catalog how many "hits" they were likely to receive. These records without linked items represented books that had not been charged out since we migrated to the NOTIS system.

When it came time to do the inventory, more than 50% of the Brooklyn Stacks collection was represented in the catalog by bibliographic records without bar codes or circulation information. To recap, there were books with no bar codes or item records because they had not circulated since at least 1982, the year we began bar coding; there were books that had bar codes but lacked corresponding item records from our CLSI system because file capacity constraints prevented them from being included in the unlinked item file; and finally, there were books with bar codes, but although the corresponding item records from our CLSI system existed, they were not linked to their bibliographic records. We also discovered other mistakes in the copy and item information, which will be described later.

A method was required that could identify not only books that were missing, but also books on the shelves that were not in the NOTIS catalog, and books that did not have item records attached to the bibliographic records. This method also needed to identify and facilitate a cleanup of the mistakes connected to the copy and item information. Two possible methods were rejected: the traditional one in which the shelflist card is matched to the book on the shelf, and the use of portable bar code scanners (Allen 1998). The former would not identify books on shelves that were absent from the NOTIS catalog, while the latter would be extremely time-consuming.¹ Neither would provide the opportunity to do a cleanup.

Method

The head of library systems proposed the idea of discharging (checking in) every book on the shelves from the location we wanted inventoried. This would have the effect of changing the "last use" date, which is the only date affected by a return (see figure 1); in other systems, incidentally, it may be necessary to charge and discharge the book to achieve the same effect. This method had the advantage of solving the problems we would encounter beyond the missing items. It would reveal the books on the shelves that were *not* in the online catalog, which the shelflist method could not do; it would also reveal mistakes in the records that could not be addressed by using portable bar code scanners. In addition, examining the online record for every book in this collection afforded a once-in-a-lifetime opportunity to do a cleanup. Beyond merely finding books that were not in the online catalog and books that lacked circulation data, we would also find tape-loaded item records linked to the wrong bibliographic record, circulation information attached to the wrong copy, and copies with no call number.

The value of this idea was in its sheer simplicity. Students would actually be able to perform most of the inventory, and they would know when a book had to be "bounced" to a professional.

Because the CUNY+ catalog contains the records for all nineteen institutions and is managed centrally, one step absolutely critical to the success of this project could not be done in-house. After every book was discharged, the programming manager of the CUNY Office of Library Systems would have to create a detailed report. This would be an exceptions list comprising every item in CUNY+ from Brooklyn Stacks that did *not* have a last use date of 1998 (the year the inventory began) or later. These would be the missing books. However, this would not be a record of every book that had been lost. Missing items that were still in the paper shelflist but had never made it into the online catalog would not be discovered. Nonetheless, that was deemed acceptable because the aim was to correct the records in our *online* catalog.

In fact, there would have to be two "lost" lists. The first would identify missing books without item records; these would be titles that had not circulated since 1987, the year of migration to NOTIS, or earlier. The second would identify missing books with item records.

Reports Required from the CUNY Office of Library Systems

Report 1—No Item Record Attached to the Bibliographic Record (figure 2). (If these books had been on the shelf, an item record would have been created as part of the inventory process.)

- This would be a list of records for books in the location Brooklyn Stacks with Library of Congress call numbers that were already cataloged, as opposed to records for books that were on order.
- The elements to be included were (1) call number, the method by which the report was sorted; (2) NOTIS record number; (3) OCLC number, taken from the *first* 035 field; (4) author's last name; (5) title; (6) date; and (7) publisher. The latter four were for reordering purposes, as was the arrangement.

Report 2—Linked Item List (Item Records Linked to the Bibliographic Record) (figure 3).

- This would be a list of linked item records that had a last use date occurring *before* 1998, with two exceptions: (1) Any book charged to a patron was omitted from this report, regardless of how long ago the book had been borrowed, because its title did not appear as available in the catalog. Moreover, on such items it was in the best interest of the library not to delete these records since it would wipe out the circulation history, thereby making it impossible to exact fines and other penalties. (2) Any book with a “creation date” of 1998 or later (the inventory period) would also be omitted from this report because in NOTIS a newly cataloged book would automatically have a creation date but not a last use date. If this group was not

omitted, those newly catalogued books that were not borrowed during the inventory period would show up on the missing list (see figure 4).

Alternatively, the cataloging unit could discharge every new book at the point of cataloging.

- This report would require the same arrangement and specifications as Report 1, except that the NOTIS number would be extended to the item level.

Preparation and Budget

The administration of the college appropriated \$250,000 above our normal budget for this endeavor. The staffing costs were \$227,000; the remainder went for equipment and supplies. The inventory would take one year to perform. There were a total of twenty-eight hourly employees, made up of computer operators and shelvers, as well as three full-time computer operators. The total number of hours devoted to this project by the part-time workers was 19,515, which included hours spent on deleting records from the local catalog. The cost of the hourly employees was \$161,000, the average pay being \$8.25 per hour. The three full-time workers cost a total of \$66,000. This budget did not include my time, that of the evening circulation supervisor, or that of the CUNY programming manager. During any given period, there were as many as nine computer operators and a supervisor.

The hours of operation were 9 A.M. to 9 P.M., Monday through Thursday, and 9 A.M. to 1 P.M. on Fridays. These hours mirrored the normal operating hours of the library during the week. It is important to point out that when time is not a factor, rather than hiring a large pool of new workers, the same procedure can be done more economically a section at a time with existing staff.

It was obvious that tackling this kind of inventory would take a great deal of organization because it needed to be completed before we moved into temporary quarters, slated for the

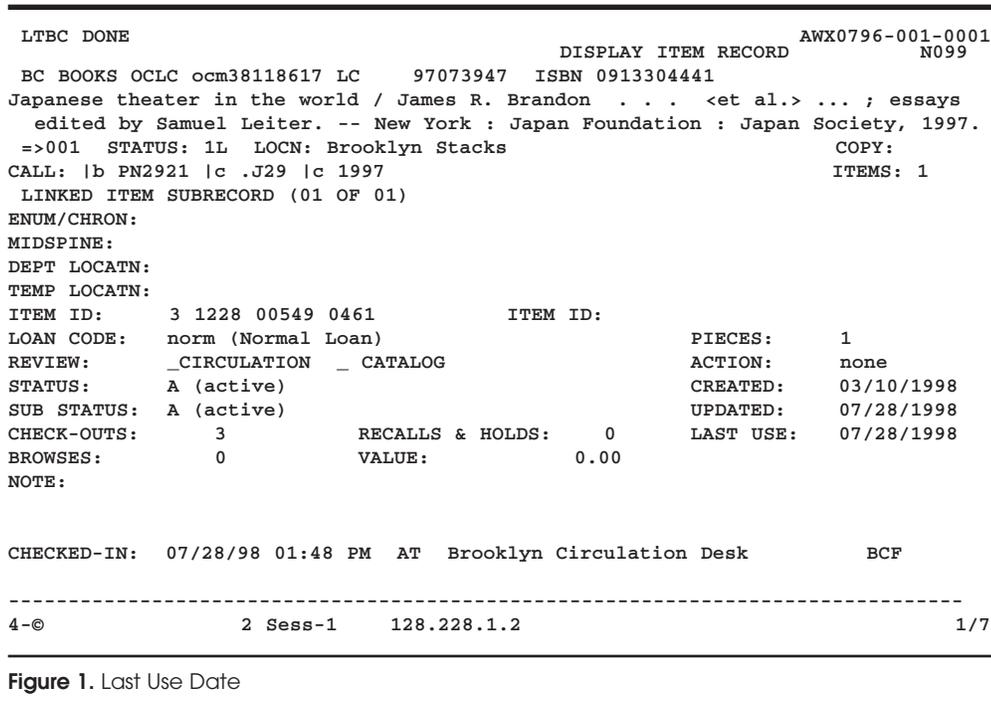


Figure 1. Last Use Date

City University of New York University Library Automation Services			
bcmn	F804.C36 K87	ACL2636	1 07835923
	Kutsche, Paul, 1927- 1981: Albuquerque : University		Ca nones, values, crisis, and
bcmn	F805.M5 G6 1969	ABQ5785	1 00052603
	Gonzalez, Nancie L. Solien, 1969: Albuquerque, University of		The Spanish-Americans of New M
bcmn	F809 .B27 1960	ACF4370	1 00479862
	Barnes, Will C. (William Cro) 1960: Tucson, University of Ari		Arizona place names. Rev. and
bcmn	F811 .C79	ACF4373	1 00479934
	Cross, Jack Lee, 1921- 1960: Tucson, University of Ari		Arizona: its people and resou
bcmn	F82.L66 C5 1970	ABQ6121	1 00126848
	Chyet, Stanley F. 1970: Detroit, Wayne State Univ		Lopez of Newport; colonial Am
bcmn	F851.7 .M7	ABU2837	1 03512741
	Miller, Tom, 1932- 1964: Seattle, The Mountaineers		Photographs by Tom Miller: the
bcmn	F852 .W28	ACE4559	1 00171625
	Warren, Sidney, 1916- 1949: New York, Macmillan co.,		Farthest frontier: the Pacifi
bcmn	F859 .G63	ACE4441	1 00606542
	Goethe, C. M. (Charles Matthi) 1949: Sacramento? 1		What's in a name? Tales, hist
bcmn	F864 .A2	ABQ4030	1 00418751
	Dillon, Richard H. 1966: New York, Coward-McCann		The legend of Grizzly Adams,
bcmn	F865 .S923	ACF7607	1 00474962
	Sutter, John Augustus, 1936: New York, G.P. Putnam's s		1803-1Sutter's own story; the life
bcmn	F868.D2 N4 1963	ACQ0912	1 03258015
	Newhall, Nancy Wynne. 1963: San Francisco : 5 Associa		Death Valley / photographs by
bcmn	F868.S5 M88 1967	ABU2883	1 00269147
	Muir, John, 1838-1914. 1967: San Francisco, Sierra Clu		Gentle wilderness; the Sierra
bcmn	F868.Y6 A36 1967	ABU2895	1 00479849
	Adams, Ansel, 1902- 1959: San Francisco, 5 Associat		Yosemite Valley. Edited by Na
bcmn	F868.Y6 M915	ABQ4443	1 01470665
	Muir, John, 1838-1914. 1948: Boston, Houghton Mifflin		Yosemite and the Sierra Nevada

Figure 2. Sample from Report 1—Missing Books without Item Records

summer of 1999. Our task was huge. Five hundred thousand books would have to be discharged. Some item records would have to be created, others linked, and some corrected. Lists of all the missing books would have to be produced and their records deleted from CUNY+. Our holdings for the corresponding records would have to be removed from the OCLC database. The copy cataloging unit would take care of recataloging for the online catalog those books that were on the shelves but not in NOTIS, a process that was not part of the inventory costs. It was agreed beforehand that the CUNY Office of Library Systems would not programmatically delete the CUNY+ records. This determination resulted from a pilot project conducted before the inventory; we had come to the conclusion that copies and volumes still on the shelves might get deleted due to the relationship between the item, vol-

ume, and copy holdings records in NOTIS. Thus, we deleted them manually in-house.

Process Inventorying the Collection

Each book was taken off the shelf, placed on a truck, and brought to one of the computer operators. While searching for the record with book in hand, the computer operator would be faced with one of several possibilities. It might be a bar coded book *with* or *without* an item record, a bar coded book whose item record had been *linked* or *unlinked*, or a book lacking a bar code *with* or *without* a bibliographic record. These possibilities and the ensuing actions are illustrated in the flowchart (figure 5). The first action on all existing item records was to discharge the book, which was done in the circulation module. When it was an unlinked item, as soon as the book was discharged, an "X" would appear in the "Catalog" field. In such a case, the operator would go into the technical services module to link the item to the bibliographic record. A book that did not have a bibliographic record was put aside for the cataloging unit after being tagged with a red dot.

The supervisors were trained in all the duties required to perform the inventory: searching by Library of Congress call number, title, and author; matching a book to a record; creating an item record; linking an unlinked item record to a bibliographic record; and discharging a book.

The supervisors, in turn, trained the computer operators. The operators added their initials to the records so that they could be identified. This facilitated periodic reviews of the operators' work, and any necessary retraining was conducted.

Correcting the Record (Quality Control)

In addition to correcting mismatched item records attached to bibliographic records (see flowchart, figure 5), operators were also trained to spot and correct multiple items attached to the same copy (for single-volume works) and to add call numbers to copies that lacked them (see figure 6). This procedure is designated "Quality Control" on the flowchart.

Teamwork

The inventory turned out to be an outstanding example of interdivisional cooperation between technical services and access services. The cataloging unit (technical services) did the inventory. The shelvers (access services) brought down the books and returned them to the shelves. The evening circulation supervisor (access services), who was essentially the inventory's "Busby Berkeley," choreographed a moving panorama of shelvers and book trucks. This enabled every computer operator to have a sufficient number of new books to work on at all times, without being bogged down in a sea of completed books.

How the Process Impacted the Readers

The evening circulation supervisor, working with the evening head shelver, devised an ingenious method for keeping track of all the books and informing the patrons of where any book in the Brooklyn Stacks collection happened to be at any given moment. Every time books were loaded onto a truck, the truck received a number. That same number was also affixed to the empty shelf. While looking for a book, if a reader approached an empty shelf, he would be instructed to jot down the number on the shelf, go to the inventory area, and glance at the trucks at each station until he found the matching number. Since the books were stacked in call number order, it did not take much effort to find the book. Once found, a book was either forfeited immediately (if it had already been discharged), discharged on the spot while the reader waited, or sent back to cataloging, where it was given same-day processing. The circulation supervisor maintained a record of the call number range of books on each truck. Throughout the inventory, all trucks were numbered consecutively, the last number being 2,627.

There was not a single complaint throughout the entire operation. In fact, there were fewer complaints during the inventory than usual because a crew of shelvers followed the books as they were returned to the shelves and shelf-read them.

City University of New York University Library Automation Services bc - Item Analysis Version: 10/27/99 Run Date: 10/27/99		PAGE	220
ACE2700-001-0002	Call number: QA37.2.A74 1985 Oclc number: 11713453	Location: bcmm	
Author: Arfken, George B. Title: Mathematical methods for physicists / George Arfken Year: 1985 Publisher: New York : Academic Press			
AFK8694-001-0001	Call number: QA37.2.B36;1988 V. Oclc number: 13560784	Location: bcmm	
Author: Bamberg, Paul G. * Title: A course in mathematics for students of physics / Year: 1988 Publisher: Cambridge ; New York : C			
AFK8694-001-0002	Call number: QA37.2.B36;1988 V. Oclc number: 13560784	Location: bcmm	
Author: Bamberg, Paul G. * Title: A course in mathematics for students of physics / Year: 1988 Publisher: Cambridge ; New York : C			
ACK0250-001-0001	Call number: QA37.2.E38 Oclc number: 05029610	Location: bcmm	
Author: Edwards, R. E. Title: A formal background to mathematics / R. E. Edward Year: 1979 Publisher: New York : Springer-Verla			
ASZ0034-001-0001	Call number: QA37.2.K6313 1991 1 Oclc number: 23384105	Location: bcmm	
Author: Koch, Helmut, Title: Introduction to classical mathematics / by Helmut Year: 1991 Publisher: Dordrecht ; Boston : Klu			
ADG7183-001-0001	Call number: QA37.2.S575;1986 Oclc number: 12103790	Location: bcmm	
Author: Smith, Douglas, Title: A transition to advanced mathematics / Douglas Sm Year: 1986 Publisher: Monterey, Calif. : Brooks			
ACC8415-001-0001	Call number: QA37.2.S87 1986 Oclc number: 12312881	Location: bcmm	
Author: Strang, Gilbert. Title: Introduction to applied mathematics / Gilbert Str Year: 1986 Publisher: Wellesley, Mass. : Welles			
AAA6617-001-0001	Call number: QA37.2.U42X;1987 Oclc number: 16502957	Location: bcmm	
Title: UMAP modules : tools for teaching 1986 / edited Year: 1987 Publisher: Arlington, MA : Consortiu			
ACV6659-001-0001	Call number: QA37.2.U42X;1988 Oclc number: 18701990	Location: bcmm	
Title: UMAP modules : tools for teaching 1987 / edited Year: 1988 Publisher: Arlington, MA : Consortiu			
ABV0523-001-0001	Call number: QA371.B79 Oclc number: 00900404	Location: bcmm	
Author: Brand, Louis, Title: Differential and difference equations, Year: 1966 Publisher: New York, Wiley 1966			
AUS4435-001-0001	Call number: QA371.F3413 1993 Oclc number: 25371334	Location: bcmm	
Author: Fedoršič, Mikhail Vasil-evič. , Title: Asymptotic analysis : linear ordinary differential Year: 1993 Publisher: Berlin ; New York : Spri			
RVN3243-001-0001	Call number: QA371.F85 1929 Oclc number: 02279245	Location: bcmm	
Author: Fry, Thornton C. Title: Elementary differential equations, by Thornton C. Year: 1929 Publisher: New York, D. Van Nostrand			
ACM7128-001-0001	Call number: QA371.H37 Oclc number: 07313216	Location: bcmm	
Author: Hassard, B. D. Title: Theory and applications of Hopf bifurcation / B.D Year: 1981 Publisher: Cambridge ; New York : C			
UA2416-001-0001	Call number: QA371.K82 1993 Oclc number: 27107133	Location: bcmm	
Author: Kuang, Yang,			

Figure 3. Sample from Report 2—Missing Books by Item Record

Other positive outcomes were similar to those experienced elsewhere (DeMiller 1991): the identification of books needing repair, of multiple copies that were set aside for possible de-selection, and of misshelved books within Brooklyn Stacks. This last benefit can help clear up situations where patrons are billed for books that are actually in the library (Stearns 1998).

Immediate Benefit

As soon as every book had been inventoried, and even before the missing lists had been produced and the records subsequently deleted, both the reference bibliographers and the circulation staff were instructed on how to interpret the catalog in light of the inventory. Thus, they could give the reader on-the-spot information that had previously required a lengthy search of the shelves. As discussed earlier in this article, the online public access catalog (OPAC)

message “check shelf” is generated when no item record is attached to that copy or volume. Previously, that message indicated three possibilities: the book was somewhere in the stacks, the book was checked out under an unlinked item record, or the book was lost. Once the books passed through this unique inventory, all books on the shelves were given linked item records, and as a result, “check shelf” took on a narrower meaning. Staff viewing this message would know for certain that the book was not there; otherwise, it would have acquired a linked item record which generates the message “not checked out” or “chkd-out, due: [date].”

Thus, whenever the message “check shelf” appeared, without any additional effort, the staff would be able to inform the reader that the book was indeed missing.

Likewise, the staff were also shown how to look for the last use date on the item record in the technical services module (this information is not available in the OPAC). If the message “not checked out” appeared but the book was not in fact on the shelves, and if the record had a pre-1998 last use date, the staff member could inform the patron that the book was gone.

Problem Solving

When conducting this kind of an inventory, one should note the following:

- In order to catch all the books to be inventoried that have been stashed somewhere else in the library, it is necessary to operate like a detective. One should request a list of all the temporary locations such as reserves, storage, and bindery. Remaining vigilant is paramount. In our case, for example, every book in the reserves collection was discharged at the onset of the inventory. Thereafter, each new influx of reserve items was discharged.

Then there are the unofficial temporary locations such as librarians’ offices. Librarians do not

```

LTBC DONE                                AWX0796-001-0001
                                           N099
BC BOOKS OCLC ocm38118617 LC      97073947  ISBN 0913304441
Japanese theater in the world / James R. Brandon ... <et al.> ... ; essays
  edited by Samuel Leiter. -- New York : Japan Foundation : Japan Society, 1997.
=>001 STATUS: 1L  LOCN: Brooklyn Stacks          COPY:
CALL: |b PN2921 |c .J29 |c 1997                ITEMS: 1
LINKED ITEM SUBRECORD (01 OF 01)
ENUM/CHRON:
MIDSPINE:
DEPT LOCATN:
TEMP LOCATN:
ITEM ID:      3 1228 00549 0461          ITEM ID:
LOAN CODE:    norm (Normal Loan)        PIECES:      1
REVIEW:       _ CIRCULATION _ CATALOG    ACTION:       none
STATUS:       A (active)                 CREATED:      03/10/1998
SUB STATUS:   A (active)                 UPDATED:      none
CHECK-OUTS:   0                          RECALLS & HOLDS: 0
BROWSES:      0                          VALUE:        0.00
NO
TE:

-----
4-@                2 Sess-1          128.228.1.2                1/7

```

Figure 4. New Book—Cataloged but Not Discharged

always subscribe to the belief that they actually have to check out a book if it is not taken out of the building.

Books that readers have taken off the shelves and left on the tables have to be discharged as well. At Brooklyn College, they are called the “sweep” books. Twice daily, they are swept up in order to be reshelved. Every day during the inventory, these books were delivered to the inventory area so that they could be discharged before being returned to the shelves.

- It is a good idea to periodically set aside specific books so as to prevent them from being discharged, using them as samples to test the programs that are being developed.
- If the library is planning to migrate to a new system, this is the ideal time to do the inventory. When ushering in a new system, it is undesirable to bring over incorrect or incomplete data. No one can predict, after all, what the new system will do with such information or whether errors will be easily corrected. Besides, the negative PR that might result could cause an additional problem.
- To avoid skipping any books, an easy way for an operator to keep track of each book that he is discharging is to turn it on its fore edge after the task is executed. However, from a preservation standpoint, remaining in this position would not be good for the life of the

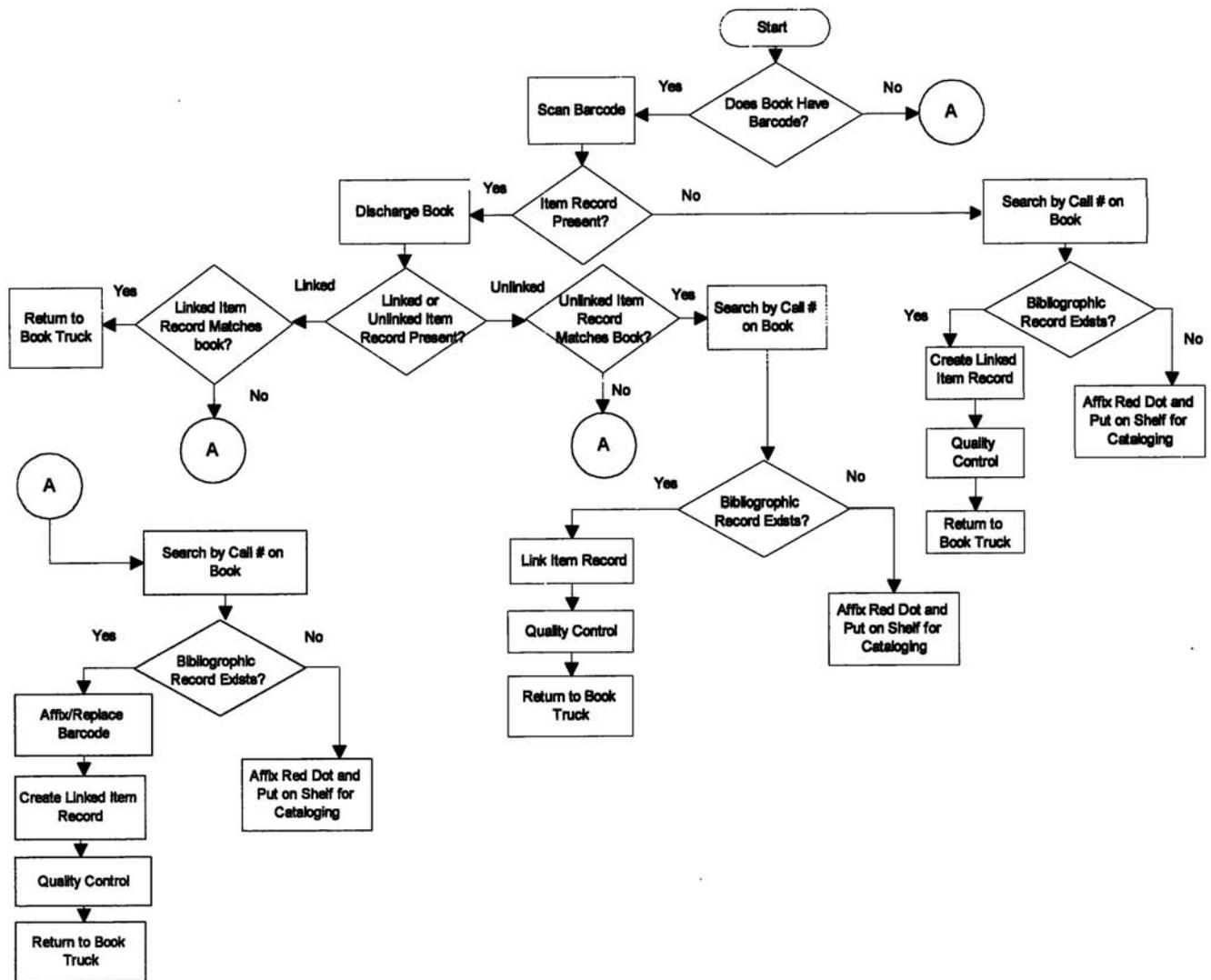


Figure 5. Inventory Procedure Flow Chart

book. Our books were never in this position for more than a few hours before being righted by the shelvers, who then returned them to the shelves.

- The computer operators will need instruction on how to read the bibliographic record to identify multivolume works. In our library, they set them aside for the two copy catalogers who were moonlighting as inventory operators. They created volume holdings when necessary.
- If the books without bibliographic records are set aside for later processing, it is helpful to distinguish them with a colored dot, as we did. Even so, they should be segregated because both temporary and permanent dots will fall off, especially if the public

has contact with them. Many of our old books had such a patina of fingerprints that nothing would stick, so we squirreled away the “red dot” books in technical services, where they would not be confused with books already in the online catalog.

- In NOTIS, a newly cataloged book will have a creation date but not a last use date. The item report of exceptions will have to omit all books with a creation date during the inventory period so that the cataloging unit will not have to discharge every new book.
- If retrospective conversion must occur concurrently with the removal of holdings from OCLC, special care should be taken. Ideally, the Dewey books should be

LTBC DONE	AYT8924	
	NOTIS COPY HOLDINGS	N001
BC BOOKS OCLC ocm42365897		
Acute leukaemia in children. -- London : Leukaemia Research Fund, 1997.		
(Information booklets and leaflets / Leukaemia Research Fund)		
STATUS h DT 12/17/2001 AD none		
NOTES		
001 1L CN a bcmn b RJ215 c . / 26 c 1997		d 12/17/01
NOTES		
ITEM RECORDS 2		
002 1L CN a bcmn		d 12/17/01
NOTES		
003 0A CN		d 12/19/01
NOTES		
004 0A CN		d 12/19/01
NOTES		
005 0A CN		d 12/19/01
NOTES		
006 0A CN		d 12/19/01
NOTES		
007 0A CN		d 12/19/01
NOTES		

4--	2 Sess-1	128.228.1.2	1/7
-----	----------	-------------	-----

Figure 6. Quality Control—Incorrectly Linked; No Call Number

new to the online catalog, but in actuality they are often duplicates of books in the LC collection. The danger is that one hand might be adding the copy to the local catalog while the other is deleting its holdings from OCLC. If this process is done manually, the person deleting from OCLC can check the creation date of each of these copies in the local catalog.

- If the inventoried collection has many old books, it might be desirable to have each book pass through the entrance sensor to check for the absence of tattle tape after each truck has been completed.

Conclusion

Once the reports were generated, we learned that 41,000 books (8%) of the Brooklyn Stacks collection were missing. The records for these books have since been deleted. Twenty-eight thousand of these were from report 1, the books that had no item record attached to the bibliographic record. While we did not have reason to believe it was a mishelving problem (Van Gemert 1996), there is no basis to conclude that the entire 41,000 had been stolen. During the 1970s, a large number of books had been inadvertently deselected from the satellite library without going through the withdrawal process, and as a result, their records remained in the catalog. This is a somewhat

comforting thought; one would normally expect that the missing items are the most used (Brazier and Reynolds 1997). The bibliographic records for 8,000 “red dot” (previously cataloged) books (1.6% of this collection) were identified and input into CUNY+. Every LC classified book in the Brooklyn Stacks collection that lacked an item record acquired one.

Even though the inventory process allowed us to fulfill all of our goals, this is not to imply that we encountered no difficulties. During our examination of the first version of report 2, the linked item list, it became clear that some books had been overlooked, so we found it necessary to redo portions of the inventory. This predicament had occurred due to human error. The lesson to be learned here is that while a cadre of inexperienced personnel might be cost-effective, supervision is essential throughout every aspect of the project.

Either the traditional shelflist method or the portable bar code scanner method would have been more efficient had we wanted merely to identify missing items. Given our circumstances, however, there is no doubt that the method we chose was superior. The paper shelflist method could not possibly reveal the 8,000 items on the shelves that had been cataloged previously but were not present in the online catalog. The portable scanner method presupposes that virtually every book has a bar code that is connected to the full cataloging record. Not only was that not the case for us, but even worse, many bar coded books had lost their item records during migration to NOTIS. This factor, plus the lapses in quality control, and the books for which there were no catalog entries in NOTIS, meant that a large number of books on the shelves would need to be retrieved for special handling. Such a requirement would offset any advantage gained by utilizing portable bar code readers in the stacks.

I was recently asked what feedback the library was receiving from such a monumental undertaking. I answered proudly, “The best feedback of all—none.” At the time of this writing, we have been in a closed stack environment in temporary quarters for more than two and a half years. The president of the college has informed me that he has not received one complaint about the library. This reaction (or more specifically, this lack of reaction) cannot be attributed solely to the accuracy of the catalog, but it surely must play a part. Additionally, the statistics kept by the circulation unit since the paging operation began provide a 98% hit rate even though some books had been mishelved during the move. Now that all the records have full circulation information, we can do maintenance inventories with portable bar code readers (Webb 1994).

We have been told by the CUNY Office of Library Systems that we have the cleanest records in the City University of New York. All our hard work was worthwhile.

Notes

- Workers would have had to make two trips to the same shelves: one to scan in the bar codes and another to remove the books that had to be added to the NOTIS catalog after each bar code exceptions report was generated. Parenthetically, if the portable bar code method is selected, it will be greatly facilitated by placing the bar code on the outside of the book rather than on the inside back cover.

Works Cited

- Allen, Virginia. 1998. Automated library inventory using bar codes. In *International Association of Aquatic and Marine Sciences Libraries and Information Centers Conference 23d: 1997*. Charleston, S.C.: IAMSLIC.
- Bowdoin, Sally. 1997. Interviewed by author. Professor Bowdoin, Head of Serials, Brooklyn College, provided information on the origin of the unlinked item records.
- Brazier, Hugh, and Brian Reynolds. 1997. What's missing? Lessons from a library stock-check. *An Leabharlann*, 2d ser., 13, no. 1: 13–21.
- Brooklyn College Library. 2001. History of the Midwood Campus. Accessed March 21, 2002, http://academic.brooklyn.cuny.edu/library/divisions_services/special_collections/virtual_exhibits/midwood_campus/index.html.
- DeMiller, Anna L. 1991. Survey of a research collection with preservation, assessment, weeding, and inventory objectives. In *Kaleidoscope: Proceedings of the Research Forum, Mountain Plains Library Association Academic Library Section*, Phoenix, Ariz., Oct. 29–Nov. 2, 1991. S.l.: The Association.
- Iskenderian, Marguerite. 1997. Interviewed by author. Professor Iskenderian, Music Cataloger, Brooklyn College, advised the author of the decisions made at the time the library became an OCLC member.
- Stearns, Melissa. 1998. Going to the shelves: Annual inventory. *Mississippi Libraries* 62, no. 4: 81–82.
- Van Gemert, Edward V. 1996. Where have all the “lost” books gone? *College and Research Library News*, no. 9 (Oct.): 581–83.
- Webb, Donna R. 1994. Sidebar 3: Inventory Minnesota State University System Project for Automated Library Systems inventory module. *Library Hi-Tech* 12, no. 1: 21–22.
- Yu, Emma Lee. 1997. Interviewed by author. Professor Yu, Cataloger, Brooklyn College, provided history of the 1975 inventory.

Book Reviews

Margaret Rohdy, Editor

Seymour Lubetzky: Writings on the Classical Art of Cataloging. By Seymour Lubetzky; comps. and eds. Elaine Svenonius and Dorothy McGarry. Englewood, Colo.: Libraries Unlimited, 2001. 443p. \$67.50 (ISBN 1-56308-932-7) LC 2001-038429.

When Seymour Lubetzky came to the Library of Congress in 1943, he was given the assignment of studying the cataloging rules to see if they could be made simpler and more logical. His approach to this assignment was to spend years in the stacks at the Library of Congress studying the history of cataloging in order to try to understand the principles implicit in the rules—the reasons for the rules or the purposes they were intended to serve. In this search for principles, he discovered the writings of Antonio Panizzi and Charles A. Cutter, which he found so useful that he spent the rest of his career quoting them and directing students to them. In his own work, he was able to take the principles first expressed by Panizzi and Cutter and suggest ways to bring the rules back to these principles, making them simpler and more logical—and therefore easier to teach, understand, apply, and “explain to inquiring readers and searching administrators” (144–45).

Now we are living in an age in which the presence of full text on the Internet has made it tempting to believe that we can do without the expense and complexity of creating surrogates for these documents in the form of metadata created by well-educated and experienced catalogers following well-designed cataloging rules and standards. Elaine Svenonius and Dorothy McGarry have performed a service of inestimable value to the cre-

ators and users of metadata, now and in the future, by publishing in this book a number of Lubetzky’s writings that had near-print or gray literature status until now—my own copies of many of these writings are photocopies of typescripts handed down from cataloging teacher to cataloging student. Now that these are all in published form, one hopes they will be immediately and widely read and will not sit on the shelf waiting for a new Lubetzky to point out that the reasons for the rules are more valid now than ever.

There are several other reasons why the publication of this book is an important event in the metadata world. One is that the origin of much of the *Anglo-American Cataloging Rules*, 2d ed. (AACR2) can be found in these writings. Here is the first suggestion of the theory of separate bibliographic identities, leading eventually to the AACR2 practice of entering one person under more than one pseudonym. Explained here is the AACR principle that a change of the name of a corporate body is a change of identity. AACR2’s entry of persons under the names by which they are commonly known, rather than under the fullest form, and entry of corporate bodies directly under their own names, rather than the earlier practice of entering them under place originate with Lubetzky as does the practice of successive entry serials. The origin of the practice of successive entry of serials can be found here. The idea that conditions of authorship of a work should be analyzed without regard for the “character of the work or the medium containing it” in order to determine how it should be identified in the catalog is pure Lubetzky and explains why

the rules in AACR2 from chapter 21 on are usually not limited to particular types of publication (with notable exceptions for art works, musical works, sound recordings, and certain legal and religious publications).

Another reason that the publication of this book is an important event is that the origin of the concepts of *work*, *expression*, and *manifestation* in Functional Requirements for Bibliographic Records (1998) can be found in Lubetzky’s important distinctions between *book* and *work*.

And finally, the publication of this book is important because these writings contain much good advice that still has not been taken. Lubetzky’s explanations as well as many explications of the nature of the bibliographic universe and of the pitfalls certain to be encountered in any attempt to control it are still relevant to the design of any system for providing access to metadata. For example, Lubetzky’s warning that “[t]he title, while quick and convenient, is not a very reliable guide” (34) should be pondered by those system designers who offer a title search to users seeking a known work, instead of a search that allows them to search using both author and title words, matching both parts of that search against authority records to catch variation in either the author’s name or the title. The quote, by the way, is from his early article “Titles: Fifth Column of the Catalog.” “Different editions should be cataloged separately, each on a different entry, but different issues of a given edition should be listed together on the same entry” (42) is an early endorsement of the practice that we now would call expression-based cataloging, as opposed to the current practice of issue-based (or

manifestation-based) cataloging that often leads to dozens of records in OCLC for the same expression of the same work, especially in the humanities. And Lubetzky's sound advice about corporate authorship, taken in AACR1, was rescinded in AACR2, leading to many more works of corporate authors being identified by title alone.

For all these reasons, this is a book that should be read by all librarians, information scientists, system designers, experts in informatics, knowledge engineers, and anyone else who ever creates or uses metadata, helps others use and interpret metadata, or designs systems for searching and displaying metadata. As Lubetzky says, "Those who are still longing for a code of rules which could be applied by a beginning cataloger without the exercise of judgment are looking backwards to a time which has gone by" (149).—*Martha M. Yee (myee@ucla.edu), University of California at Los Angeles, Film and Television Archive.*

Works Cited

- Anglo-American Cataloguing Rules*. 2d ed. 1988 rev. Eds. Michael Gorman and Paul W. Winkler. Ottawa: Canadian Library Association; Chicago: American Library Association.
- Functional Requirements for Bibliographic Records: Final Report*. 1998. IFLA Study Group on the Functional Requirements for Bibliographic Records. München: K. G. Saur. Also found (March 5, 2002) at www.ifla.org/VII/s13/frbr/frbr.pdf.

Subject Analysis in Online Catalogs.

By Hope A. Olson and John J. Boll. 2d ed. Englewood, Colo., Libraries Unlimited, 2001. 333p. (ISBN 1-56308-800-2) LC 2001-029828.

This book is an expanded and updated edition of a 1991 book by Rao Aluri, Alasdair Kemp, and John Boll. In his review of the first edition, Wellisch (1991) predicted the need for a new edition within a few years.

Indeed the new edition reveals just how much growth there has been in the relevant literature in the past ten years. Much of this work is focused on maximizing the value of established subject access tools in the environment of online catalogs. In this regard, it demonstrates a body of innovative thought following the pioneering work of Cochrane (1985, 1986).

The format—part encyclopedia, part commentary, and in general a guide for the perplexed—is like a medieval compendium. The topic could be quite narrow, but instead the authors have viewed it as the complex intersection of two larger topics. The fundamental principles and basic structures of both subject analysis and online catalogs are presented with concision and agility. Topics are interrelated and linked by internal references within the text. Both the core literature of each topic and the most recent research are cited extensively. The coverage of gray press research reports, many available on the Internet, is impressive.

The analysis is original and serves to relate the discrete concepts to the overall theme. The authors also identify current problems and prospects for future developments and research. Both the challenges to and the opportunities for improving online subject access are described in detail. The role of authority control and the online use of classification are the two main issues.

Bringing together various types of databases through the catalog presents a new need for vocabulary control across multiple files as well as within the catalog. Enhancements such as adding tables of contents to bibliographic records reflect a demand for higher levels of exhaustiveness and specificity in searching the catalog. The authors caution that "it is not clear that the ramifications of doing so have been carefully weighed to ensure that more positive (higher recall) than negative (lower precision) results are produced"

(320–21). The impact of vocabulary control freed from linear file structures is presented in the context of combining sophisticated retrieval techniques.

The revival of interest in classification is viewed with enthusiasm. Again the authors warn that "even in an electronic environment order and linearity, and at least some of the traditional principles of classification, cannot be ignored" (186). These principles provide a counterbalance to the scattering effect of expanded indexing of controlled and uncontrolled search terms. The potential for online classification in support of hierarchical and lateral browsing is immense. The widespread, if clumsy, use of classification by Internet search engines is witness to this potential. The development of more flexible and transparent system-user interfaces suggested by Boll and Olson will be a critical step forward.

The literature of bibliographic control seems to draw readers largely among technical services librarians. This book has as much or more to offer to public service librarians and library system designers. It is a gold mine of bibliographic instruction strategies. Library automation vendors could gain a competitive edge by studying some of its chapters. In at least one MLS program, this book will be used as the text for a subject analysis course. Although not intended per se to prepare catalogers, it provides an excellent basis for a course of value to information professionals in all fields.—*J. Bradford Young (jbyoung@pobox.upenn.edu), University of Pennsylvania Libraries, Philadelphia.*

Works Cited

- Cochrane, Pauline Atherton. 1985. Classification as a users' tool in online public access catalogs. In *Redesign of Catalogs and Indexes for Improved Online Subject Access: Selected Papers of Pauline A. Cochrane*. Phoenix, Ariz.: Oryx Press. First published in

Proceedings of the 4th International Study Conference on Classification Research, 1982. (Frankfurt: Indeks-Verlag, 1982).

- Cochrane, Pauline Atherton. 1986. *Improving LCSH for Use in Online Catalogs: Exercises for Self-help with a Selection of Background Readings.* Littleton, Colo.: Libraries Unlimited.
- Wellisch, Hans. 1991. Review of *Subject Analysis in Online Catalogs*, by Rao Aluri, D. Alasdair Kemp, and John Boll. *Library Resources & Technical Services* 35: 479–80.

The Invisible Web: Uncovering Information Sources Search Engines Can't See. By Chris Sherman and Gary Price. Medford, N.J.: Information Today, 2001. 439p. \$29.95 (ISBN 1-55938-510-3) LC 2001-028818.

The popularity of Internet search engines belies their weaknesses as tools for information discovery; using one is like panning for gold, combing through a scoopful of wet sand for a few valuable nuggets. While the unmanageable size and bewildering variety of search results are obvious to—and taken in stride by—most search-engine users, the fact that the large preponderance of Web-based content is beyond the reach of the Internet search services is much less apparent and unsuspected by many. This is the phenomenon of the invisible Web, which experienced searchers and webmasters Chris Sherman and Gary Price seek not only to make us more aware of, but also to equip us to use effectively. In a nutshell, this book sets out to do three things: (1) define the invisible Web; (2) show the would-be searcher how to access the information it contains; (3) present a “starter’s kit” collection of representative invisible Web resources.

For the most part, Sherman and Price handle the first two tasks very well indeed. Chapters 1 through 8 are a model of pedagogical technique. The

authors start by providing basic concepts and necessary background, continue by defining the invisible Web in relation to the Web and the Internet in general, show how a searcher can tell whether a site is likely to be “visible” or “invisible,” list the types of information likely to be found on the invisible Web, and use case studies to show how the invisible Web can be successfully exploited by novice as well as experienced searchers. Organization is clear and logical, categorical analysis is used to good advantage, strategic repetition reinforces key concepts, and well-placed sidebars clarify important terms and puncture myths as they surface. The style is relaxed and informal while maintaining a high level of focus.

However, this relaxed approach is not without a few problems. One is that the authors’ definition of “invisible Web” is not as clear and consistent as it could have been. The definition encapsulated in the book’s subtitle is lucid enough and is admirably fleshed out in the discussion of the various types of invisibility. Resources likely to be invisible to search engines include databases, sites that require registration, and the deeper pages of an unusually large site. The authors muddy the waters, however, when in some places they refer to “opaque” pages that are indexable by search engines, but not indexed, as part of the invisible Web but elsewhere exclude them from that category. Definition of terms sometimes sacrifices clarity and rigor in favor of casual readability, most egregiously in the glossary definition of “precision,” which lacks a key final phrase from the definition used in the body of the text. Even with such minor defects, however, the first third of the book provides an effective and engaging guide to Web searching in general as well as the relationship between invisible Web resources—especially databases—and the Web as a whole.

The remaining two-thirds of the text consists of a classified, annotated

directory of about 1000 selected invisible Web resources. It comprises 18 subject or genre categories and 127 subcategories and appears to be a thin and very eclectic sampling of the types of “invisible” resources being discussed. Annotations—either composed or cut and pasted from a self-description—are provided for most resources; the rest are attached to annotated entries as “related resources.” Wherever possible, URLs are provided for both a site’s search form and home page.

This directory is also on the Web at www.invisible-web.net, but it is disappointing to find, some five months after the book’s August 2001 printing date, that despite the authors’ stated intention to expand and update their directory, this has not happened so far. The online content does not go beyond that of the book, and there are more than a few broken links. Even so, this directory is well enough organized and offers a sufficient variety of quality sites to give any novice searcher valuable learning experience with the invisible Web. While it seems questionable to devote so much print to a reference feature that is sure to become dated quickly, and which is both available and much more usable online, the directory does advance the purposes of the book by serving as an artist’s rendering of what the invisible Web has to offer. The “Web guide” section of the book, it should also be noted, contains descriptions of several directories and specialized search tools the authors have found handy for locating invisible Web resources; this is one of the book’s more useful features.

Part how-to manual, part compilation of sites, part background briefing paper, *The Invisible Web* is both more and less than what it appears to be. Its stated objective is to provide a “map” of the vast reservoirs of Web-based content inaccessible to the major search engines. While this metaphor may have energized the authors’

efforts and will likely pique the reader's interest, it may also raise expectations this book cannot fulfill. Nevertheless, if what was meant was a conceptual map of the invisible Web—and the authors do not state this—the effort has succeeded handsomely. The wealth of useful concepts, distinctions, and examples, and the carefully organized way they are presented, make Sherman's and Price's book a remarkably valuable field guide for anyone seeking content on the Web.—*Gregory Wool (gwool@iastate.edu), Iowa State University, Ames.*

Maps and Related Cartographic Materials Cataloging, Classification, and Bibliographic Control. Eds. Paige G. Andrew and Mary Lynette Larsgaard. New York: Haworth Information Pr., 1999. 487p. (ISBN 0-7890-0813-0) LC 99-51487.

If the editors had not included a section called “Those That Got Away” (xvii–xviii) in the introduction to *Maps and Related Cartographic Materials Cataloging, Classification and Bibliographic Control*, I would not have realized that some parts of the cartographic cataloging universe are not included in this impressive array of chapters by prominent members of the map cataloging community. This book includes everything from the basics of map cataloging to spatial metadata to retrospective conversion.

The book is a well-organized how-to guide for cataloging diverse types of cartographic materials. After the introduction and general information about cataloging maps, chapters on related topics are grouped into sections: “Cataloging Specific Material Types,” “Handling Early Cartographic Material,” “Digital Cartographic Materials,” “Classification and Subject Access of Cartographic Materials,” “Retrospective Conversion of Collections and Quality Control,” and “Cartographic Materials in an Archival Setting.”

The editors have preserved the tone of each chapter, which gives the reader a feel for the background and experience of the authors but results in inconsistencies within the text. For instance, some chapters include the AACR2 (Anglo-American Cataloguing Rules, 2d ed.) rule number references in the text, while others put the references in endnotes. Unfortunately, the authors' original comma usages are preserved; a few authors use commas so sparingly that sentences are puzzling until their context within the paragraph is understood. I found myself mentally inserting commas into sentences such as “Whereas titles of books are usually evident from the title page maps quite often provide more than one title from which to choose” (40).

Map cataloging is at a crossroads. One of the primary manuals, *Cartographic Materials: A Manual of Interpretation for AACR2* (1982), has been out-of-print for years, though one can buy an overpriced black-and-white copy printed on demand. A long-awaited revision is imminent. Similarly, the *Map Cataloging Manual* (1991), prepared by the Geography and Map Division, Library of Congress, is being revised. Neither manual could have anticipated the World Wide Web and the explosion of digital cartographic data. Neither manual adequately covers the cataloging of early maps. In effect, *Maps and Related Cartographic Materials Cataloging, Classification and Bibliographic Control* is the only current reference manual devoted to the bibliographic description of cartographic materials. The chapters are written in an organized, simple style ideal for the beginning map cataloger. Frequent references to the two older manuals and the primary tools of AACR2 and the MARC21 (Machine Readable Cataloging) format allow readers to look up the original citations and judge for themselves whether they accept the authors' interpretation of the best way to catalog the cartographic resources. Clearly, the authors

intend to provide a pragmatic and detailed supplement to the primary cataloging tools.

Even an experienced map cataloger will appreciate the chapters on the cataloging of special formats that they encounter infrequently. I recently referred to “Cataloging Aerial Photographs and Other Remote-Sensing Materials” in cataloging several photo-mosaic indexes of areas in Tennessee. Unsure of how to interpret some of the numbers on the photo-mosaics, I consulted *Maps and Related Cartographic Materials Cataloging, Classification and Bibliographic Control* and quickly found the information needed to determine date and scale. As is typical with other chapters in the “Cataloging Specific Material Types” section, the authors include background information on the map format and discussion of what is significant for cataloging. The text is accompanied by numerous photo-reproductions of the maps and examples of MARC records. The authors cite many references for additional information.

The number and quality of illustrations and catalog record examples vary from chapter to chapter. Those on aerial photographs and on early printed maps are among the best in providing illustrations and accompanying MARC examples. “Cataloging Geologic Sections” is invaluable for its illustrations of different types of geologic sections and explanations of cataloging technique, but contains not one example of a complete bibliographic record. Pictures of map series or atlases would add little to two of the most clearly written chapters in the compilation, “Cataloging Map Series and Serials” and “Cataloging the Contemporary Printed Atlas,” though all of the “how-to” chapters would be enhanced by full-level MARC catalog records, accompanied when practical by illustrations of the resources. In some cases, the catalog records seem to have been an afterthought because they illustrate pre-

AACR2 cataloging rules or have MARC-coding errors. Though the chapters on metadata include comprehensive lists of citations for further exploration, the entire “Digital Cartographic Materials” section disappoints in its lack of *any* complete bibliographic records, whether in Dublin Core, MARC, or some other format.

The chapters on early maps and map archives, digital cartographic materials, and retrospective conversion projects are a good introduction to areas of growing importance. Early map catalogers as well as those on the cutting edge of geospatial metadata description are challenged by the changing nature of practice, standards, and reference sources in these fields. “Cataloging Early Printed Maps” is particularly good for reconciling con-

flicting information among manuals.

A comprehensive reference work needs a comprehensive index. Unfortunately, the index is one of the few weak points of the book. Cross-references are limited, making it challenging to find terms. I could not find “Raster” in the index anywhere, and “Resource Description Framework” was listed only as a subcategory under “Metadata.” When this work proves to be so useful that an updated edition is published, I hope the editors will include both a comprehensive index and a combined bibliography of all the cataloging resources cited in the bibliographies of the separate chapters.

None but the most fanatic map cataloger will read *Maps and Related Cartographic Materials Cataloging, Classification and Bibliographic Con-*

trol from cover to cover. Its strength is as a handy reference tool for specific areas of cartographic description and access and for planning that long-overdue retrospective conversion project.—*Kay G. Johnson (johnsonk@utk.edu), University of Tennessee Library, Knoxville.*

Works Cited

- Library of Congress. 1991. Geography and Map Division. *Map Cataloging Manual*. Washington, D.C.: Library of Congress, Cataloging Distribution Service.
- Cartographic Materials: A Manual of Interpretation for AACR2*. 1982. Prepared by the Anglo-American Cataloguing Committee for Cartographic Materials. Chicago: American Library Association.