Information Technology and Libraries

Volume 15 Number 4 ISSN 0730-9295

December 1996

Information Technology and Libraries (ISSN 0730-9295) is published quarterly in March, June, September, and December by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Library and Information Technology Association, a division of the American Library Association. Subscription price, $25, is included in membership dues. Nonmembers may subscribe for $50 per year in the U.S.; $55 in Canada, Mexico, Spain, and other PUAS countries; $60 in other foreign countries. Single copies, $15. Periodical-class postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to Information Technology and Libraries, 50 E. Huron St., Chicago, IL 60611.

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 8 of the March 1996 issue. Manuscripts of articles, communications, and news items should be addressed to the editor: James J. Kopp, Library Director, Wilson W. Clark Memorial Library, University of Portland, P.O. Box 83017, 5000 N. Willamette Blvd., Portland, OR 97283-0017. Copies of books for review should be addressed to: Barbara Kemp, ITAL Book Reviews, Dewey Graduate Library, University of Albany, SUNY, 135 Western Ave., Albany, NY 12222. Copies of software for review should be addressed to: Tom Jevic, ITAL Software Reviews, University of Illinois at Chicago, Main Library, 861 S. Morgan, Chicago, IL 60607.

As a matter of policy, *Information Technology and Libraries,* as the scholarly organ of LITA, does not review LITA publications. Notice of new publications from LITA will generally be found in the "Other Recent Receipts" column following reviews.

PRODUCTION: ALA Production Services (David Epstein, Bruce Frausto; Gwen Ihnat, Christine Squires, Leslie Stella, and Donavan Vicha), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Information Technology and Libraries

Volume 15, Number 4 December 1996 ISSN 0730-9295

Cover Design by Jim Lange. Interior Design by Dianne M. Rooney

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk,* and *Science Abstracts Publications.* Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, New Search,* and *Social Sciences Citation Index.* Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

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. ∞

EDITORIAL

Paul Evan Peters: Friend, Mentor, Prophet

James J. Kopp

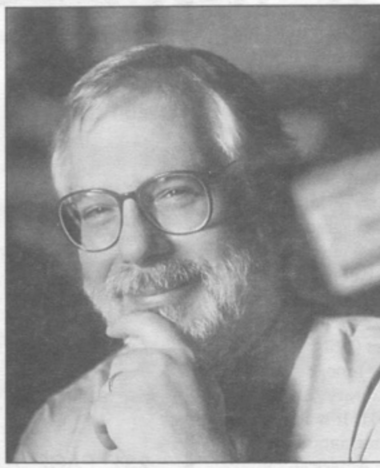
Paul Evan Peters was a friend. I imagine a large portion of the readers of this journal considered Paul a friend, as do many others beyond the readership of this publication. Most individuals who encountered Paul Peters probably considered him a friend, even if that encounter was a brief one; even if it was not a direct, personal one but perhaps consisted of seeing Paul at the podium at one of his many professional presentations, in the lounge at one conference or another, or even virtually through e-mail or through some of his other online contributions. Paul Evan Peters epitomized friendliness.

Paul was a very close friend to our profession. The bonds of this friendship were his contributions within the institutions with which he was associated, including Columbia University, the New York Public Library, and the Coalition for Networked Information; within the profession as a whole; and especially to the Library and Information Technology Association, which he served as its president in 1991–92 and in so many ways.

I first met Paul when I interviewed at Columbia University in 1984. I had never been to Columbia previously, in fact had never been to New York City, so you can imagine the anxiety and nervousness of this experience. Paul was on the search committee, and when I sat down with that group he made some typical calming remark, and all of a sudden it became one of the most relaxing and pleasant experiences I have had in such circumstances. We instantly had a rapport, something that many, many others have experienced with Paul. I figured at the time it was some sort of bearded bonding, especially as our whiskers were mutually showing more and more a "touch of gray."

Although I did not work directly with Paul nor did I interact with him a great deal in our time together at Columbia, I came to consider him a mentor, probably something that he would abhor and cherish at the same time. The wisdom and guidance I obtained from Paul concerned not only library systems, about which I learned much, but also some of those other things that made Paul Paul: his sense of humor in a stressful situation, the calmness amidst some systems crisis, a world view of libraries, and how to have a good time when the work was done (and occasionally before). I am sorry I did not have more time to work with Paul while I was at Columbia. I am even more sorry that I did not have an opportunity years later to tell him how much he was a mentor to me. Somehow though, I think he knew. Then and now I think that when I grow up I want to be like Paul Evan Peters. Maybe it's a bearded bonding thing.

There's another bearded connection in all of this, too. In reflecting on Paul and his contributions, I came to think of another fellow who has played a significant role in my life over the past two decades. Many of you might not know the name, but those who know me have heard of me speak of Edward Bellamy, the "hero" of my master's thesis and a key figure in my doctoral dissertation. Bellamy was a nineteenth-century journalist turned reformer whose utopian novel, *Looking Backward 2000–1887*, became one of the most popular books of its time and one of the most influential books in American history, with such individuals as Thorsten Veblen, Thomas Dewey, Frank Lloyd Wright, and the two presidential Roosevelts owing something to Bellamy's work and beliefs. The utopian world of the year 2000 depicted in Bellamy's novel relied a great deal on technological advancements as well as changes in the human outlook and condition (and even a few comments on improvements in libraries). That in itself makes me think of Paul, but there are other threads of connection. Bellamy looked to the future but did so with an eye on the past and the present. I believe Paul did that also, and I find that one of the activities that he undertook with much interest and characteristic zeal was very indicative of this. "The Global Library" exhibit at New York Public, of which he was the curator and which many of you probably experienced at Annual Conference in June (or viewed via the World Wide Web), combined the importance of the historical as well as the promise and potential of the future. That combination is at the heart of this exhibit. It also was at the heart of the work and beliefs of Edward Bellamy and, I believe, at the heart of Paul Peters.



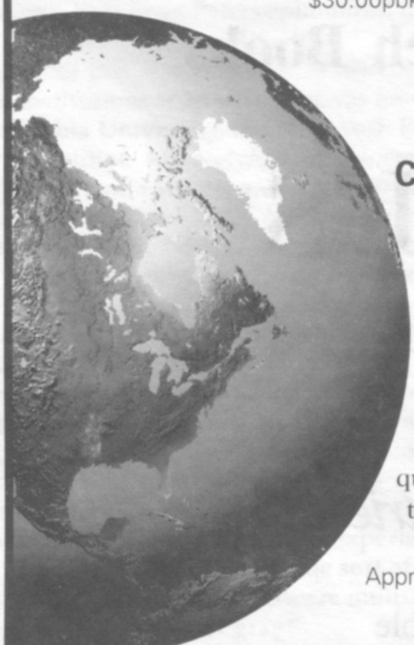
Paul Evan Peters

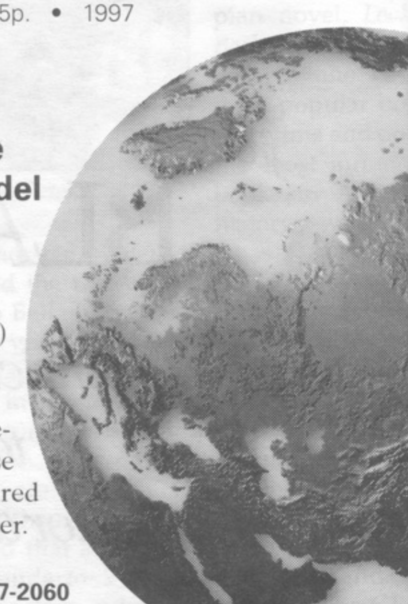
Edward Bellamy was considered a prophet of his time. So, too, do I feel that Paul Evan Peters might be considered a prophet of his time and of his profession. When Bellamy died in 1898 (like Paul at age 48), a tribute written to Bellamy in the *Arena* stated that:

> When a prophet dies, he does not go away; he disappears. The real presence of the departing seer remains behind and expresses itself evermore in his works.

This eulogy was prophetic itself, in that Bellamy's presence remained behind and expressed itself in his works. In a similar way, Paul's presence remains behind and will express itself evermore in his works, in his friends, and in those to whom he was a mentor.

Here's to you, Paul.





The Design of a Relational Database for Large-Scale Bibliographic Retrieval

Rebecca Green

A fully normalized relational bibliographic database promises relief from the update, insertion, and deletion anomalies that plague bibliographic databases using (US)MARC formats internally. The conceptual design of a full-scale bibliographic database (including bibliographic, authority, holdings, and classification data) is presented, based on entity-relationship modeling. This design translates easily into a logical relational design. The treatment of format integration and the differentiation between the intellectual and bibliographic levels of description and between collective and individual levels of description are discussed. Unfortunately, the complexities of bibliographic data result in a tension between the semantic integrity of the relational approach and the inefficiencies of normalization and decomposition. Compromise approaches to the dilemma are outlined.

According to Date, "It is undeniable that the relational approach represents the dominant trend in the marketplace today, and that the 'relational model' . . . is the single most important development in the entire history of the database field."[1] Despite the popularity of this data model in the larger database world, large-scale bibliographic databases have typically retained data structures based on the complex, but unitary MARC record. While some degree of relationality is implied—for example, in the use of the USMARC bibliographic, authority, holdings, and classification formats—only limited direct interrecord linkage occurs;[2] the bibliographic database world is still considerably more record-oriented than relation-oriented. (Reference to MARC in this article is often USMARC-specific. Nevertheless, the larger conclusions presented here should hold also for other MARC implementations.)

Given the general effectiveness of relational databases, it is not surprising that the idea of redesigning bibliographic databases according to relational database principles has cropped up from time to time in the literature.[3] Starting from a similar premise, a seminar conducted in the College of Library and Information Services (CLIS) at the University of Maryland in fall 1994[4] undertook to establish the basic logical design of a large-scale bibliographic database using the entity-relationship (ER) model, with a view to the eventual conversion of the ER-based conceptual schemas into a relational database. The results of that undertaking are reported and discussed in this article.

Background

MARC and Database Design

In the bibliographic world, the MARC family constitutes an unrivaled standard. Designed for the transfer of bibliographic data in machine-readable form, specifically on magnetic tape, the MARC formats are, first and foremost, communication formats. For lack of suitable alternatives, they have also been used as storage formats. In the standard three-level database architecture,[5] which distinguishes among internal schemas (physical data storage structures), conceptual schemas (logical, community-wide views of the data), and external schemas (user views of the data, especially in terms of output reports or screen displays), the MARC formats generally correspond most closely to external schemas.

This three-level database architecture supports the ideal of data independence, the capacity to make changes in one level or schema of the database without having to replicate changes in other levels of the database.[6] On the one hand, the use of MARC records as a communications format on the input and/or output side dictates neither the logical view of a bibliographic database nor its internal storage structure. On the other hand, the internal storage structure must be MARC-compatible, so that data coming in from a MARC record can be transformed to be consistent with data already in the database and data in the database can be transformed into a MARC record for output; Llorens and Trénor introduce just such a system.[7] Indeed, in order for a system to be fully MARC-compatible, this transformation process should be reversible: if a MARC input record is transformed into the physical data structures of a given bibliographic database and that data is transformed back into a MARC output record, the output record should match the input record; there should be no "loss of data or content designation."[8]

While these transformations are more easily accomplished if the storage structure resembles the MARC format, they do not actually require a MARC-like internal schema. A decision on the type of file structure to be adopted for internal storage should instead take into account all the functions required of the database. If a hybridization of several file structure types is called for, as will likely be the case, it should be forged with knowledge of the frequency of all necessary database processes as well as of the costs and benefits associated with various data modeling options.

Rebecca Green is Associate Professor, College of Library and Information Services, University of Maryland, College Park; e-mail rgreen@umd5.umd.edu.

In contrast to a unitary MARC-like storage format, Gorman has long envisioned bibliographic retrieval systems based on "*linked packages* of information."[9] More recently, he has called for "simple records with complex connections" to replace "the MARC record . . . a complex record that has no connection with any other record,"[10] a call consistent with relational bibliographic databases. Attig notes that while authority relationships, holdings relationships, and such bibliographic relationships as analytics, translations, and revisions are "crucial" for descriptive cataloging and lie "at the heart of the collocating function of the catalog," they are "very difficult to communicate."[11] Accepting Gorman's vision for the sake of discussion, Attig—who personally favors the creation of "a single self-contained [super]record with all related records treated as subrecords"—argues that if bibliographic data is to be transmitted from one relational system to another, there is no guarantee that entire sets of related records will be transferred or that equivalent linkages will be established in the receiving system. These problems, though real, are, however, not unique to bibliographic databases, but come with the territory where data transfer involving interrecord relationality is concerned. Thus Attig's concern does not bear directly on Gorman's vision of simple data packages with complex interconnections as a conceptual data view or as a data storage strategy.

Modeling of Bibliographic Data

Traditional fixed-length, flat-file record formats face several stiff challenges with bibliographic data. For one thing, bibliographic data elements are highly variable in length: A title, for example, may be a single word—only a handful of characters long—or may be dozens of words long. For another, of the many bibliographic data elements that may apply to bibliographic items (the USMARC bibliographic format now includes hundreds of tags, each of which may encompass several distinct data elements), only a reasonably small percentage are likely to apply to any specific item. Lastly, some data elements applying to a given item may occur multiple times, for example, multiple statements of responsibility, multiple places of publication, multiple notes. The basic approach of the USMARC format to these challenges consists of gathering together all data relating to a single bibliographic item into one bibliographic record and imposing control over the variability within that data through a highly structured directory segment, a map to the variable-length fields within the record.

Deciding on appropriate methods for storing such complex data calls for examining the attributes of bibliographic entities as well as the interrelationships among them: We need to understand the semantics of the data. Of extant semantic data modeling techniques, the most widely accepted is entity-relationship (ER) modeling.[12] This view of data modeling recognizes a primary distinction between *entities* (fundamental things in the environment that is being modeled) and *relationships* (associations between or among entities); both entities and relationships can have *attributes* (properties). For instance, a physical book is an entity with attributes such as title, material type, and dimensions. Similarly, a library patron is an entity with attributes such as name and social security number. Circulation, however, is a relationship between the book and the patron; due date is an attribute of that relationship.

A systematic series of steps translates entity-relationship data models into relational data models. If the ER modeling and this translation are both done properly, the result will be a fully normalized relational database, one free of insertion, deletion, and update anomalies.[13] ER modeling is thus a natural prelude to the design of a sound relational bibliographic database.

Within the library and information studies world, ER modeling has been applied fruitfully to both AACR2 and USMARC. Fidel and Crandall use ER modeling to show that AACR2 mixes rules of different types, resulting in "confusion between the internal/external and conceptual levels of design and a poorly organized presentation of cataloging rules."[14] On the basis of his analysis, Leazer suggests that the USMARC bibliographic format includes data elements that are attributes of such nonbibliographic entities as persons, corporate bodies, and places; recognizing distinct entity types "would allow the relationship between entities to detail the function that one entity serves in relation to another."[15] This modeling technique has thus proven itself applicable to the bibliographic context.

Methodology

In accordance with the perceptions expressed above, the CLIS seminar began with the premise that the design of a relational bibliographic database would be more theoretically sound than the current traditional flat-file design. This database design process can be broken down into six steps: (1) requirements analysis, (2) conceptual database design, (3) choice of a database management system, (4) data mapping into a logical database design, (5) physical database design, and (6) database system application.[16] Of these, the seminar tackled only requirements analysis, conceptual database design (using entity-relationship modeling), and logical (relational) database design.

Using resources common in the U.S. bibliographic world, the group as a whole applied the principles of ER modeling to the data elements addressed in chapter 1 (general rules for description), chapter 2 (books, pamphlets, and printed sheets), chapter 12 (serials), chapter 13 (analysis), chapter 21 (choice of access points), and chapter 26 (references) of AACR2[17] and the USMARC formats for bibliographic, authority, holdings, and classification data.[18] Because our output was in a state of continuous flux, we contented ourselves with a textual rendition of our ER modeling rather than capturing our work in a standard entity-relationship diagram; it is in this state that the conceptual schema for the database is presented in the appendix. In addition to the chapters indicated above, the appendix also incorporates individual students' work on chapter 4 (manuscripts) and chapter 6 (sound recordings) of AACR2. Throughout this phase our emphasis was not on exhaustivity but on deciding how to treat the most important as well as the most problematic data elements. Nevertheless, the final product is relatively extensive.

Requirements Analysis

Conducting a full-blown requirements analysis for a specific information system is a mammoth task. When the system to be designed is general and theoretical, the task can easily expand, thereby becoming impossibly huge. Rather than attempt to inventory all potential bibliographic needs, which would have resulted in an undesirable expansion of the requirements analysis task, we predetermined a fixed set of needs the system would address. This kept the task manageably small yet still reasonably comprehensive.

Our final statement of the functions of a bibliographic database—the product of our requirements analysis—is shown in figure 1. It recognizes four overall purposes: identification of bibliographic items, selection of appropriate bibliographic items, showing the availability of bibliographic items, and showing relationships between bibliographic items (the statement of relationships is based on Tillett[19]). We owe an obvious debt to Cutter's "objects": (1) to enable a user to find material whose author, title, or subject is known; (2) to show what is available in the collection by a specific author, on a specific subject, or in a specific kind of literature; and (3) to assist in the choice of a document.[20] We were also influenced by Hagler's summary of the overall uses to which bibliographic data elements are put: "to identify a particular document uniquely . . . , to show how two or more documents are associated . . . , and to provide the basis for access points."[21]

Database Design

A full decade ago, Crawford, Becker, and Ogilvie applied relational modeling to bibliographic data.[22] The result of

Identification

Verify that a bibliographic unit exists.

Verify information about a bibliographic unit so as to make accurate and distinguishing reference to it.

Ascertain whether a bibliographic description matches a bibliographic unit.

Ascertain whether two bibliographic units are the same.

(a) bibliographic entity ("book")

(b) intellectual entity ("work")

Selection

Determine the (absolute) appropriateness to a user's situation of

(a) a bibliographic entity.

(b) an intellectual entity.

Determine the (relative) appropriateness to a user's situation of two

(a) bibliographic entities.

(b) intellectual entities.

Availability

Determine the physical location of a bibliographic unit.

Determine (restrictions on) availability (e.g., hours facility is open, document delivery delay, required user authorization) of a bibliographic unit for

(a) local use.

(b) loan (including interlibrary loan).

(c) purchase.

Relationships

Determine the relationship between a work and any of its bibliographic manifestations, e.g., copies, photocopies, reprints ("equivalence").

Determine the relationship between a work and any works produced by modifying it, e.g., revisions, translations, abstracts, dramatizations, parodies ("derivative").

Determine the relationship between a work and a work that describes it, e.g., reviews, criticisms, annotations, commentaries ("descriptive").

Determine the relationship between an intellectual or bibliographic entity and its components, e.g., extracts, anthology selections, journal articles ("whole-part").

Determine the relationship between a bibliographic unit and another bibliographic unit that accompanies and either complements or augments it, e.g., parts of kits, concordances ("accompanying").

Determine the relationship between a work and a predecessor work, e.g., sequels, parts of numbered series, successive serial titles ("sequential").

Collocate works or bibliographic units that have some characteristic (e.g., author, subject, language, date of publication) in common ("shared characteristic").

Figure 1
Functions of a Large-Scale Bibliographic Database

their work was a fully normalized set of relations, a logical database design for a relational bibliographic database. The work of the seminar differs from their work in several respects, all related to scope. We treated bibliographic data, authority data, holdings data, and classification data; they treated only bibliographic data. We treated the entire bibliographic domain in a general way while also treating several specific formats (e.g., monographs, serials) in detail; they treated only journal articles. We treated both functional relationships (dependencies) between attributes and relationships between entities; they treated only the functional relationships between attributes.

Llorens and Trénor's recent sketch of a relational bibliographic database bears a much closer resemblance to our effort;[23] unfortunately, the limited detail in which their database design is presented precludes extensive comparison. Furthermore, the database design they present is a working physical design that represents certain compromises. Although their logical design is not recoverable from that physical design, it seems fair to suggest it was not as extensive as ours. Despite these differences, the analysis presented here and the conclusions drawn there appear compatible.

Our conceptual database design began by building on our requirements analysis. For each function recognized there, we identified the data elements required for its performance. For example, verifying that a bibliographic item exists implies knowing at least its title and publication date. Additional data elements were identified based on Barry's study of user-centered relevance criteria;[24] still others were introduced gradually throughout the seminar on the basis of elements in the various USMARC formats (we did not include USMARC elements pertaining to the record itself). We then took this basic inventory of bibliographic data elements and asked of each element, "Of what entity or relationship is this an attribute?" A title, for example, is an attribute of a bibliographic item. In contrast, a due date is an attribute of the relationship between a patron and a bibliographic item checked out to the patron. Through this process we identified a basic set of entities and relationships for a general bibliographic database and associated with each entity and relationship its attributes. To grossly oversimplify our results, our basic data model at this stage involved four major entity types: bibliographic items, bodies (both individual and collective), concepts/subjects, and names. Basic relationships between these entity types included the following:

BODY <produces> BIBLIOGRAPHIC ITEM
BIBLIOGRAPHIC ITEM <held in> BODY
BIBLIOGRAPHIC ITEM <loaned to> BODY
BIBLIOGRAPHIC ITEM <conveys> CONCEPT/SUBJECT
BODY <has> NAME
CONCEPT/SUBJECT <has> NAME

Discussion

Several aspects of our logical design warrant extended discussion: (1) the treatment of bibliographic entities (a quartet of entity types lumped together as "bibliographic item" above); (2) AACR2 compatibility; (3) the strategy for handling variety across bibliographic formats; and (4) the tension between normalization and efficiency, which leads to the consideration of (5) alternative approaches.

Bibliographic Entities

The distinction between books and works, between bibliographical units and literary units, is well accepted, but reveals only the tip of the iceberg. O'Neill and Vizine-Goetz, limiting themselves to monographs, identify five hierarchical levels of bibliographic entity. First, there is the book, "an individual physical manifestation of a bibliographic entity"; above the book comes the printing, "a set of books printed at one time or printed at different times containing no more than slight variations"; then comes the edition, "a set of printings produced from substantially the same type image"; one step further comes the text, "a set of editions with the same content"; lastly at the top of the hierarchy is the work, "a set of related texts with a common origin and content."[25] Each entity is related to one and only one entity at all higher levels of the hierarchy, but may be, and is typically, associated with many entities at lower levels.

Although the naming of their bibliographic entities has a decidedly monographic bias, O'Neill and Vizine-Goetz claim universal applicability for their concepts across all types of bibliographic materials. Our treatment of the bibliographic entity issue differs from theirs in several minor, but not wholly insignificant, respects. First, we decided not to be concerned with the entity level represented by printings, since information at this level is generally not recorded (see AACR2's rule 2.2D1). Second, we attempted to devise names for the levels that could generalize across material types but still retain some degree of communicativeness. On the physical side, we dubbed the lowest level, representing an individual physical entity, a BIBLIOGRAPHIC COPY; the set of bibliographic copies produced under essentially the same circumstances we called a BIBLIOGRAPHIC SET. On the literary side, we referred to a more or less fixed version (e.g., edition, translation, paraphrase, condensation) of an intellectual unit as a

WORK and to an abstract, virtual intellectual product (e.g., a text in all its versions) as a UNIFORM WORK, borrowing from its most prominent attribute, uniform title.[26] Third, while retaining a one-to-many/hierarchical relationship from UNIFORM WORK to WORK:

WORK (1,1) <is a version of> UNIFORM WORK (1,n)[27]

and from BIBLIOGRAPHIC SET to BIBLIOGRAPHIC COPY:

BIBLIOGRAPHIC COPY (1,1) <is an instance of> BIBLIOGRAPHIC SET (0,n),

we identified the relationship between WORK and BIBLIOGRAPHIC SET as being many-to-many:

WORK (1,n) <appears in> BIBLIOGRAPHIC SET (1,n).

On the one hand, a BIBLIOGRAPHIC SET may be associated with many WORKs, as when an anthology contains many short stories. On the other hand, a WORK may also be associated with many BIBLIOGRAPHIC SETs, as when a short story appears in several different anthologies. Fourth, because some relationships involve the bibliographic item on several of these levels, we also recognized an entity type called BIBLIOGRAPHIC UNIT, the union of UNIFORM WORK, WORK, and BIBLIOGRAPHIC SET.[28]

The presence of the BIBLIOGRAPHIC UNIT "superentity" allowed us to collapse all other connections among bibliographic items into two relationships:

BIBLIOGRAPHIC UNIT/1 (0,n) <is related to> BIBLIOGRAPHIC UNIT/2 (0,n), and

BIBLIOGRAPHIC UNIT/1 (0,n) <has part> BIBLIOGRAPHIC UNIT/2 (0,n).

Indeed, the <has part> relationship could have been subsumed within the more general <is related to> relationship were it not for the fact that the attributes associated with the two relationships are not easily merged. Examples in the appendix indicate how broadly both of these relationships can be interpreted.

AACR2 Compatibility

In listing attributes we were torn between designing a bibliographic database "from scratch" and devising an AACR2-compatible storage structure. This indecision resulted in the dual desire to recognize discrete, manipulatable data items, but also to retain AACR2-based data elements, some of which defy systematic breakdown.[29] This duality of purpose is manifest in the inclusion, where needed, of both types of attributes;[30] this occurs, for example, within the BIBLIOGRAPHIC SET entity, which corresponds best with the bibliographic entity described by AACR2. In the appendix, which displays the output of our conceptual design, attributes that are preceded by a plus sign (+) are discrete data elements; although they are included conceptually within AACR2-based descriptive statements, they cannot easily be derived from the statements. Nearby (usually preceding) data elements with similar names are the corresponding AACR2-based statements.

Bibliographic Formats

The recent adoption of format integration within the USMARC world is a simplifying step, removing the necessity of defining which fields can occur with which formats. This approach is not just a matter of streamlining the USMARC definitional enterprise. It is made almost necessary by the continuing emergence of new formats, especially those that are hybrids of existing formats (e.g., electronic serials).

Within the context of the USMARC record, there are no inefficiencies associated with format integration. If a field applies to an item, it will be used; if it doesn't, it won't. But the relational database context is another matter. Here it does not make sense to list for BIBLIOGRAPHIC SET all the attributes that might apply, knowing that for any specific item most of them would not be applicable and would be marked with a NULL "value," thus wasting a good deal of storage space. Elmasri and Navathe recommend instead that if an attribute does not apply to the majority of entity class members, a separate relation should be established.[31] However, the number of attributes not applying to most BIBLIOGRAPHIC SET entities renders this alternative equally nonsensical for our situation.

Although it too is not an optimal solution, we have adopted yet a third approach to the format integration problem. We have recognized a number of material types operating along different parameters (although we have not worked out the accompanying attribute list for most of them). For example, differentiated by sensory system used, we have visual materials, aural materials, and tactile materials; differentiated by symbol system used, we have language materials, printed music, cartographic materials, and digitally encoded files. Associated with each material type are the attributes appropriate to it. Then, adopting the same basic philosophy that spawned format integration, we recognize that a specific bibliographic item may be several material types simultaneously. The typical monograph would be both visual material and language material; an electronic journal would be both a digitally encoded file and a serial publication. The use of the same key (bib_unit_id) for all records pertaining to the same bibliographic item permits their integration on the external level.

Normalization and Efficiency

The normalization of data relations is undertaken to avoid data "anomalies." Anomaly types identified include update anomalies, insertion anomalies, and deletion anomalies. Update anomalies occur when data is structured in such a way that a given "fact" may be stored multiple times in the database. Not only does this mean wasting storage space, but it also opens up the possibility of inconsistency, since when a data value changes, it may be updated in some, but not all, database locations. Of course, the data need not be updated to be inconsistent; the data could be entered inconsistently in the first place. Leazer notes a "large degree of redundancy in the USMARC formats," including repetition of essentially the same information in different parts of the record.[32]

Insertion anomalies arise when an independent "fact" has been made dependent on other data, so that it cannot be recorded in the database unless and until the other data is recorded. Deletion anomalies share the same condition of treating as dependent data what is in fact independent, but arise in situations where other such data has been recorded and is subsequently being deleted, resulting in the deletion of the independent, and still valid, data. Treating independent data as dependent data occurs within AACR2 when publisher address is treated as being dependent on items published. Consequently, it is not possible in a USMARC-based bibliographic database to record a publisher address until data is entered for some bibliographic item associated with the publisher.[33] Subsequently, if all items published by the publisher are withdrawn from the collection and the related data deleted, the independent fact of the publisher's address will also be deleted.

Such anomalies can be avoided by first identifying carefully what each data element in the database is dependent on; this directly affects what it is made an attribute of. Such careful examination is reflected in our placement of attributes with entities and relationships. The process is completed by applying a series of steps that map (extended) entity-relationship conceptual schemas into relational database schemas.[34] Because this process is largely mechanical, we have not produced a final list of data relations, but the general shape of that output is easy enough to discern. At a minimum there will be a separate relation (which will ultimately manifest itself as a separate file) for each entity class with single-valued attributes other than the key; for each multivalued (simplex or composite) attribute; for each many-to-many binary relationship; and for each relationship of degree higher than 2 (e.g., a ternary relationship). This would result in a database with approximately one hundred relations. Few databases are of this complexity, but then few contexts are as complex as the bibliographic world. It should be noted further that what has been modeled extends beyond even the largest of integrated bibliographic systems.

Normalization works by decomposition, grouping into separate relations those data elements that are dependent on the same key data elements. The decomposition enables the avoidance of data anomalies, but also disperses related data elements. For example, a BIBLIOGRAPHIC SET may have multiple parallel titles, multiple other title information statements, and multiple statements of responsibility. Since they vary independently of one another, three separate relations would be set up, even though all of them are facts about the same entity class. Given the relational design of this database, it would be necessary to access dozens of relations to pull together the full bibliographic information associated with a specific physical item,[35] an extremely common transaction in a bibliographic database. This means in turn considerable time spent in input operations, which are notoriously slow in comparison with internal processing operations. Translating between USMARC data and this relational storage structure would be similarly I/O-bound. The cleanness of the relational design seems almost immaterial in the face of these inefficiencies.

Alternate Approaches

This rejection of a full relational approach does not, however, vindicate a MARC-like storage format for bibliographic data. Although the nonrelational design of the MARC record makes all the joining of data required by the relational model unnecessary, and although its directory-driven structure keeps the record relatively compact, the semantic soundness of the MARC record is compromised by its susceptibility to data anomalies.

Can a bibliographic database be both semantically sound and efficient? Our understanding of semantic integrity in database design is for all intents and purposes defined by the principles of normalization. Given this understanding, we are forced to compromise: Adherence to the principles of normalization requires a level of decomposition of bibliographic data that results in inefficient processing, particularly when retrieving data.

Since for bibliographic databases the ideals of semantic integrity and efficiency are at least in partial opposition, we can imagine characterizing various bibliographic database designs according to their positions along a continuum between these ideals. The two designs already considered stake out positions toward the extremes of the continuum. A fully normalized relational design retains semantic integrity at the expense of efficiency, while the use of a MARC-based storage format contrariwise retains efficiency at the expense of semantic integrity.

Two other positions, one favoring the efficiency pole of the continuum and the other favoring the semantic integrity pole of the continuum, also warrant consideration. Since both efficiency and semantic integrity are important, middle positions on the continuum, which preserve some of both qualities, may be preferred over positions toward the extremes.

The first of these alternatives retains USMARC formats as the basis of storage records, recognizing only a few different record types: bibliographic, authority, holdings, and classification. Rather than making each record more or less self-contained, however, relational links are established between the records, for example, from a bibliographic record to every authority record that pertains to one of its access points, to every classification record that pertains to a call number assigned to it, as well as to every holdings record associated with it. This approach is exemplified by the Western Library Network, which from its outset has stored, in the place of 1xx and 7xx values in the bibliographic record, pointers to the corresponding authority records. This has resulted in a level of data integrity that other bibliographic utilities have lacked and with only minimal loss of efficiency. Most of the redundancy of the USMARC record remains, however.

The second alternative retains the basic design of our entity-relationship analysis, but moves to a different basic data structure than the relation. By definition, every attribute value in a relation is single-valued. Consequently, multivalued attributes generate additional relations. Almost one-third of the relations required to translate our conceptual schemas into a fully normalized relational design result from multivalued attributes. In comparison to a relational design, the use of a data structure that accommodates multiple values for a single attribute would reduce the inefficiencies associated with massive decomposition, but without significant deterioration of data integrity. Object-oriented databases permit the inclusion of multiple values for the attribute of an entity through their use of collection or bulk type constructors (e.g., sets, bags, lists). An added advantage is their implementation of hierarchical inheritance. Furthermore, rapprochement between object-oriented and relational database approaches within the foreseeable future is being held out as a tantalizing promise. The temptation to see object-oriented databases as a panacea is strong, but the truth of the matter remains that this design is still a compromise—perhaps the best compromise we can achieve, but a compromise nonetheless.

Conclusion

The natural complexities of bibliographic data are numerous:

- Many attributes are highly variable in length.
- The set of data elements needed to describe a given bibliographic unit is somewhat idiosyncratic, varying from one bibliographic unit to the next.
- Some data elements occur multiple times.
- The network of relationships between and among entities is intricate.

The first and last of these sources of complexity are handled easily enough within the relational data model: Variable-length fields can be accommodated with SQL's *varchar* data type, while the proliferation of relationships can be accommodated by the inclusion of foreign keys as a linking device between entities. To the contrary, the existence of many data elements that do not occur exactly one time per bibliographic unit results in massive overdecomposition of relations.

The abandonment of progress toward a large-scale relational bibliographic database does not signal overall failure for our undertaking. The conceptual design presented in the appendix is not data-model specific and stands as a continued challenge to bibliographic database design. Meanwhile, the shortcomings detected by Fidel and Crandall in AACR2 and by Leazer in the USMARC formats persist. A thorough review of the data elements needed for computerized bibliographic description and access is called for, possibly occasioning modifications in the conceptual design presented here. (The most likely result would be simplification of the BIBLIOGRAPHIC SET entity; this reduction of complexity might well be offset by the further addition of entities, attributes, and/or relationships.) In any case, the relationality inherent in bibliographic data needs better and more widespread treatment than it now receives. Investigation into bibliographic database design should continue, with especial attention given to the possible development of object-oriented bibliographic databases.

Acknowledgments

Alex Schwarzmann and Wei Ding proved particularly helpful in ironing out design details after the seminar had concluded. They are also the sources of the detailed treatment of sound recordings and manuscripts, respectively.

References and Notes

1. C. J. Date, *An Introduction to Database Systems*, 6th ed. (Reading, Mass.: Addison-Wesley, 1995), 22.

2. Walt Crawford, *MARC for Library Use: Understanding Integrated USMARC*, 2d ed. (Boston: G. K. Hall, 1989), 176–86.

3. Robert G. Crawford, H. Sue Becker, and Janet E. Ogilvie, "A Relational Bibliographic Database," *Canadian Journal of Information Science* 9, no. 1 (1984): 21–28; Terrence A. Brooks and Esther G. Bierbaum, "Database Management Systems: New Homes for Migrating Bibliographic Records," *Library and Information Science Research* 9, no. 4 (1987): 327–39; Roy Chang, "dBASE, Relational Data Models, and MARC Records," *Technical Services Quarterly* 10, no. 2 (1992): 81–91; J. Llorens and A. Trénor, "MARC and Relational Databases," *The Electronic Library* 11, no. 2 (1993): 93–97; Georgina Verity, "Relational Database Management Systems and Open Systems Used in the Development of Oracle Libraries," *Program* 27, no. 1 (1993): 73–82.

4. Participants in LBSC 772, "Seminar in the Organization of Knowledge," included CLIS M.L.S. and Ph.D. students Joseph Berlin, Wei Ding, Michael Schanno, Alexander Schwarzmann, Paula Strain, Deborah Wassertzug, and Aleksandra Zajackowski.

5. See, for example, Ramez Elmasri and Shamkant B. Navathe, *Fundamentals of Database Systems*, 2d ed. (Redwood City, Calif.: Benjamin/Cummings, 1994), 26–27.

6. Ibid., 28.

7. Llorens and Trénor, "MARC and Relational Databases," 95–96.

8. W. Crawford, *MARC for Library Use: Understanding Integrated USMARC*, 263–65.

9. Michael Gorman, "Cataloging and the New Technologies," in *Foundations of Cataloging: A Sourcebook*, ed. Michael Carpenter and Elaine Svenonius (Englewood, Colo.: Libraries Unlimited, 1985), 239–52 [244].

10. Gorman is quoted in John C. Attig, "Descriptive Cataloging Rules and Machine-Readable Record Structures: Some Directions for Parallel Development," in *The Conceptual Foundations of Descriptive Cataloging*, ed. Elaine Svenonius (San Diego: Academic Press, 1989), 135–48 [141].

11. Ibid., 141–43.

12. Elmasri and Navathe, *Fundamentals of Database Systems*, 40–63, give a thorough overview of the basic features of this technique.

13. Jerry V. Caswell, "Normalization: A Method for Structured File Design," *Information Technology and Libraries* 3, no. 3 (1984): 293–97.

14. Raya Fidel and Michael Crandall, "The AACR2 as a Design Schema for Bibliographic Databases," *Library Quarterly* 58, no. 2 (1988): 123–42 [123].

15. Gregory H. Leazer, "An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Schema for the USMARC Formats," *Library Resources & Technical Services* 36, no. 2 (1992): 189–208 [201].

16. Elmasri and Navathe, *Fundamentals of Database Systems*, 451–52.

17. *Anglo-American Cataloging Rules*, 2d ed., 1988 revision (Chicago: ALA, 1988).

18. *USMARC Format for Authority Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1993); *USMARC Format for Bibliographic Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1994); *USMARC Format for Classification Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1991); *USMARC Format for Holdings Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1989).

19. Barbara B. Tillett, "A Taxonomy of Bibliographic Relationships," *Library Resources & Technical Services* 35, no. 2 (1991): 150–58.

20. Charles A. Cutter, *Rules for a Dictionary Catalog*, 4th ed., rewritten (Washington, D.C.: Government Printing Office, 1904), 12.

21. Ronald Hagler, *The Bibliographic Record and Information Technology*, 2d ed. (Chicago: American Library Association, 1991), 26.

22. Crawford, Becker, and Ogilvie, "A Relational Bibliographic Database," 21–28.

23. Llorens and Trénor, "MARC and Relational Databases," 96.

24. Carol L. Barry, "User-Defined Relevance Criteria: An Exploratory Study," *Journal of the American Society for Information Science* 45, no. 3 (1994): 149–59.

25. Edward T. O'Neill and Diane Vizine-Goetz, "Bibliographic Relationships: Implications for the Function of the Catalog," in *The Conceptual Foundations of Descriptive Cataloging*, ed. Elaine Svenonius (San Diego: Academic Press, 1989), 167–79 [172–74].

26. Attempts to give strict definition or characterization to these entity types invariably led to the identification of exceptions. In the end we decided that the determination of entity type should be a function of which attributes pertain to the entity.

27. Relationships are shown in the form:
ENTITY1 (min,max) <relationship> ENTITY2 (min,max)

Min values give the minimum number of times each entity within the associated entity class ***must*** be involved in the relationship; a min value of 0 states that it is not the case that all members of the entity class must participate in the relationship, while a min value of 1 or n (meaning "many") states that it is the case that all members of the entity class must participate in the relationship. Max values give the maximum number of times an entity within the associated entity class ***may*** be involved in the relationship; a max value of 1 states that no member of the entity class may participate in the relationship more than once, while a max value of n states that members of the entity class may participate multiple times in the relationship. Min values determine an entity class' "degree of participation" in the relationship (partial or total); max values correspond to the cardinality ratios (one-to-one, one-to-many, many-to-many) of binary relationships.

28. A similar tactic established a BODY "superentity," which includes both PERSONS and CORPORATE BODIES, which in turn include, in addition to the prototypical corporate body, MEETINGS (e.g., conferences) and COLLECTIONS (i.e., document collections, of any size).

29. This is not to say that AACR2 areas and subareas cannot be broken down into elemental data units, but to suggest that

many of them cover such a broad array of data elements that it is not always possible to generate a standard AACR2 statement on the basis of the distinct data elements. In some cases (e.g., the extent of item statement within the physical description area) considerable massaging would be required to move from the level of discrete data element to the level of AACR2-compatible statement.

30. Even were the AACR2 statement fully derivable from the discrete data elements, it would make sense to include both. Since the underlying data elements are stable, the AACR2 "derivation" need only be processed once and stored, a more efficient strategy than generating it separately each time the bibliographic record is to be displayed.

31. Elmasri and Navathe, *Fundamentals of Database Systems*, 398.

32. Leazer, "An Examination of Data Elements for Bibliographic Description," 201–3.

33. Add data on a second item published by the same publisher, and should one choose to record the publisher address a second time, the potential for update anomaly is introduced.

34. Elmasri and Navathe, *Fundamentals of Database Systems*, 172–77 and 629–32, summarize these steps.

35. First come the BIBLIOGRAPHIC COPY relation and the separate relation associated with its multivalued notes attribute. Next come the BIBLIOGRAPHIC SET relation and the ten relations associated with each of the ten multivalued attributes of its corresponding entity class. Then come relations representing nonfunctional and/or nonbinary relationships involving BIBLIOGRAPHIC COPY or BIBLIOGRAPHIC SET (or ultimately, BIBLIOGRAPHIC UNIT, UNIFORM WORK, or WORK).

Appendix
Entity-Relationship Analysis of a Large-Scale Bibliographic Database

Key:

(x, y)	composite attribute
x(s)	multivalued attribute
*x	key attribute
*[x]	key attribute of weak entity class, borrowed from its identifying owner
+x	atomic attribute isolated from preceding AACR2-compliant composite attribute
: x	scope note
[ex.: x]	example values

A. Entities and attributes

1. BIBLIOGRAPHIC UNIT
   - *bib_unit_id
2. UNIFORM WORK: an abstract, virtual intellectual product (e.g., a text in all its versions)
   - *[bib_unit_id]
   - uniform title
   - variant title(s)
   - parallel title(s)
   - reports-successful-procedure/methodology indicator
   - summary
3. WORK: a more-or-less fixed version (e.g., edition, translation, paraphrase, condensation) of an intellectual product
   - *[bib_unit_id]
   - (rating type, rating)(s)
   - (language role [ex.: sung, spoken, written], language)(s)
   - nature, scope, artistic form
   - intended audience
4. BIBLIOGRAPHIC SET: the set of copies produced under essentially the same circumstances
   - *[bib_unit_id]
   - (title proper, title script/alphabet)
   - alternate title

(parallel title, sequence indicator)(s)
(other title information statement, sequence indicator)(s)
(statement of responsibility, sequence indicator)(s)
edition statement
(edition parallel title, sequence indicator)(s)
(edition statement of responsibility, sequence indicator)(s)
copyright date
GMD
extent statement
+(extent quantity, extent unit)(s)
other physical details statement
dimensions statement
+container type
+(maximum height, maximum width, maximum depth)
accompanying material statement
note statement(s)
standard number statement
(identifying number type [ex.: +ISBN, +ISSN, LCCN, SuDoc, PubNo], identifying number)(s)
terms of availability statement
(honor/award type, honor/award class, date)(s)
(classification scheme, class notation)(s)
restriction(s) on use
[digital object: text, graphics, audio, etc.]

5. BIBLIOGRAPHIC COPY
   *[bib_unit_id]
   *bib_copy_id
   master copy indicator
   note statement(s)
   (classification scheme, prefix, class number, suffix [including Cutter number, year of publication, etc.], copy number)
   circulation status
   condition
   restriction(s) on use statement
   (action type, action date)(s)
6. VISUAL MATERIAL
   *[bib_unit_id]
   print size
   illustration type(s) [ex.: ill., maps, music, coats of arms]
   color indicator
7. AURAL MATERIAL
   *[bib_unit_id]
8. TACTILE MATERIAL
   *[bib_unit_id]
   Braille type
9. DYNAMIC MATERIAL
   *[bib_unit_id]
10. STATIC MATERIAL
    *[bib_unit_id]
11. LANGUAGE MATERIAL
    *[bib_unit_id]
    (table-of-contents term, table-of-contents location)(s)
    ("back-of-book" index term, "back-of-book" index location)(s)
    ISBN

12. CARTOGRAPHIC MATERIAL
*[bib_unit_id]
13. PRINTED MUSIC
*[bib_unit_id]
14. DIGITALLY ENCODED FILE
*[bib_unit_id]
15. MANUSCRIPT
*[bib_unit_id]
nature, scope, form/type of manuscript
material type
opening words
number of hands
style of handwriting
number of columns
marking of lines
number of lines
overall height of lines
illustration description
composite manuscript statement
16. THREE-DIMENSIONAL OBJECT/REALIA
*[bib_unit_id]
17. MICROFORM
*[bib_unit_id]
18. SERIAL PUBLICATION
*[bib_unit_id]
key-title
frequency
regularity
status [ex.: ongoing, suspended, complete]
has-index indicator, index statement
edition-type indicator
ISDS center/ISSN network center
type of serial
19. SERIAL ISSUE
*[bib_unit_id]
(issue indicator, issue indicator type(s) [ex.: chronological, numeric, alphabetic])
20. ORGANIZATIONAL TOOL (e.g., a thesaurus, a classification scheme)
*[bib_unit_id]
21. PERFORMANCE UNIT
*[uniform_perf_id/event_id]
22. UNIFORM PERFORMANCE: production of play, opera, ballet, etc.
*uniform_perf_id
uniform name
nature
place of premiere
date of premiere
23. PERFORMANCE/EVENT: concert; performance of play, opera, ballet, etc.; festival; studio recording session; any other event
*event_id
name/caption
nature/description
place
date
time range

number
(admission fee range(s), currency)

24. BODY
    body_id
    (address, date range)(s)
    (honor/award type, honor/award class, date)(s)

25. PERSON
    *[body_id]
    (phenomenon type(s) [ex.: birth, marriage, death], phenomenon date)

26. CORPORATE BODY
    *[body_id]

27. MEETING
    *[body_id]
    name
    number
    date
    location

28. COLLECTION
    *[body_id]
    (user type, borrowing period)(s)
    (day of week, open time, close time)(s)
    holiday(s) observed

29. GOVERNMENT BODY
    *[body_id]
    gov_body_type

30. CONCEPT
    concept_id
    hierarchical level
    scope note

31. NAME
    name_id
    type of name indicator
    explanatory note

32. PERSON NAME
    *[name_id]
    surname/entry element
    rest of name

33. CORPORATE BODY NAME
    *[name_id]
    main heading
    addition to heading

34. CONCEPT NAME (incl. subject heading)
    *[name_id]
    kind of name indicator
    concept name/main heading
    (subdivision, sequence indicator)(s)

35. CLASS NUMBER
    class_number_id
    immediate level caption
    (note type, sequence indicator, note content)(s)

36. SINGLE CLASS NUMBER
    *[class_number_id]
    notation

37. CLASS NUMBER RANGE
    *[class_number_id]
    (class number range start, class number range end)

B. Relationships

B.1. Hierarchical Relationships

1. BIBLIOGRAPHIC UNIT <is a category/subset of the union of> UNIFORM WORK, WORK, BIBLIOGRAPHIC SET
2. PERFORMANCE UNIT <is a category/subset of the union of> UNIFORM PERFORMANCE AND EVENT/PERFORMANCE

In the following specialization by sensory perception, BIBLIOGRAPHIC SET participates partially; the specialization is overlapping.

3. VISUAL MATERIAL <is a kind of> BIBLIOGRAPHIC SET
4. AURAL MATERIAL <is a kind of> BIBLIOGRAPHIC SET
5. TACTILE MATERIAL <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by occurrence in time, BIBLIOGRAPHIC SET participates partially; the specialization is overlapping.

6. DYNAMIC MATERIAL <is a kind of> BIBLIOGRAPHIC SET
7. STATIC MATERIAL <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by symbol system used, BIBLIOGRAPHIC SET participates partially; the specialization is overlapping.

8. LANGUAGE MATERIAL <is a kind of> BIBLIOGRAPHIC SET
9. CARTOGRAPHIC MATERIAL <is a kind of> BIBLIOGRAPHIC SET
10. PRINTED MUSIC <is a kind of> BIBLIOGRAPHIC SET
11. DIGITALLY ENCODED FILE <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by method of recording, BIBLIOGRAPHIC SET participates partially; the specialization is trivially disjoint.

12. MANUSCRIPT <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by dimension, BIBLIOGRAPHIC SET participates partially; the specialization is trivially disjoint.

13. THREE-DIMENSIONAL OBJECT/REALIA <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by size of image, BIBLIOGRAPHIC SET participates partially; the specialization is trivially disjoint.

14. MICROFORM <is a kind of> BIBLIOGRAPHIC SET

In the following specialization by recurrence of publication, BIBLIOGRAPHIC SET participates partially; the specialization is disjoint.

15. SERIAL <is a kind of> BIBLIOGRAPHIC SET
16. SERIAL ISSUE <is a kind of> BIBLIOGRAPHIC SET

In the following use-kind of specialization, BIBLIOGRAPHIC SET participates partially; the specialization is trivially disjoint.

17. ORGANIZATIONAL TOOL <is a kind of> BIBLIOGRAPHIC SET

In the following specialization, BODY participates totally; the specialization is disjoint.

18. PERSON <is a kind of> BODY
19. CORPORATE BODY <is a kind of> BODY

In the following specialization, CORPORATE BODY participates partially; the specialization is disjoint.

20. MEETING <is a kind of> CORPORATE BODY
21. COLLECTION <is a kind of> CORPORATE BODY
22. GOVERNMENT BODY <is a kind of> CORPORATE BODY

In the following specialization, NAME participates totally; the specialization is overlapping.

23. PERSON NAME <is a kind of> NAME
24. CORPORATE BODY NAME <is a kind of> NAME
25. CONCEPT NAME <is a kind of> NAME

In the following specialization, CLASS NUMBER participates totally; the specialization is disjoint.

26. SINGLE CLASS NUMBER <is a kind of> CLASS NUMBER
27. CLASS NUMBER RANGE <is a kind of> CLASS NUMBER

B.2. Nonhierarchical Relationships

1. WORK (1,1) <is version of> UNIFORM WORK (1,n)
2. WORK (1,n) <appears in> BIBLIOGRAPHIC SET (1,n)
3. BIBLIOGRAPHIC COPY (1,1) <is an instance of> BIBLIOGRAPHIC SET (0,n)
4. BIBLIOGRAPHIC UNIT/1 (0,n) <is related to> BIBLIOGRAPHIC UNIT/2 (0,n)
   (relationship type [ex.: edition, translation, supplement, continuation, merger, split, review, paraphrase, citation, adaptation, illustration, reproduction, arrangement, transcription, accompaniment, index, concordance, dramatization, remake, colorization, remastering], relationship modifier [ex.: sequence indicator, date, language])*
5. BIBLIOGRAPHIC UNIT/1 (0,n) <has part> BIBLIOGRAPHIC UNIT/2 (0,n)
   (type indicator [ex.: volume, issue, track, chapter, section], sequence value [ex.: 1, A, July 1995], first/middle/last indicator)(s)
6. PERFORMANCE UNIT/1 (0,n) <has part> PERFORMANCE UNIT/2 (0,n)
   (type indicator [ex.: movement, act], sequence value, first/middle/last indicator)(s)
7. PERFORMANCE UNIT (0,n) <is interpretation of> WORK (0,n)
8. PERFORMANCE UNIT (1,n) <is produced by> BODY (0,n)
   function(s) [ex.: producer, director, choreographer, stage designer, costume designer, light designer, sponsor, impresario, rehearsal coach, engineer, conductor, orchestra, chorus, narrator]
9. EVENT/PERFORMANCE (0,n) <is version of> UNIFORM PERFORMANCE (1,n)
10. BODY (0,n) <plays character in> EVENT/PERFORMANCE (0,n)
    character name
11. BIBLIOGRAPHIC SET (0,n) <records> EVENT/PERFORMANCE (1,n)
    duration
12. UNIFORM WORK (0,1) <honors> PERSON (0,n)
13. WORK (0,1) <written as thesis at> CORPORATE BODY (0,n)
    (degree name, degree date)
14. WORK (0, n) <reports> MEETING (0,n)
15. CORPORATE BODY (0,n) <sponsors> MEETING (1,n)
16. CORPORATE BODY (0,n) <has part> CORPORATE BODY (0,1)
17. BODY (0,n) <physically produces> BIBLIOGRAPHIC SET (0,n)
    (function [ex.: (hand)writing, typing, printing, publication, distribution, engineering], place, (date, date type))(s)
18. NAME (0,n) <intellectually produces> WORK (0,n)
    (function [ex.: author, compiler, editor, illustrator, reviser, translator, engraver, artistic-reproducer, arranger, transcriber, reporter, medium, performer, director, choreographer, composer, set designer], certainty of attribution indicator)(s)
19. PERSON (1,n) <has> PERSON NAME (1,n)
20. CORPORATE BODY (1,n) <has> CORPORATE BODY NAME (1,n)

*BIBLIOGRAPHIC UNIT1 is the original, older entity throughout. Many relationships will lack a value for relationship modifier.

21. PERSON NAME (0,n) <is represented as> CORPORATE BODY NAME (0,n)
22. NAME (0,n) <is related to> NAME (0,n)
    relationship type (ex.: full/acronym, later/earlier, use/used for [i.e., authorized descriptor/lead-in vocabulary])
23. NAME (0,n) <is part of> NAME CLUSTER (1,n)
24. NAME (0,n) <authorized by> ORGANIZATIONAL TOOL (0,n)
    history note
25. BIBLIOGRAPHIC UNIT (0,n) <conveys> CONCEPT (0,n)
26. CONCEPT NAME (1,n) <maps to> CONCEPT (0,n)
27. CLASS NUMBER (1,n) <maps to> CONCEPT (0,n)
28. CONCEPT NAME (0,n) <assigned to> BIBLIOGRAPHIC UNIT (0,n)
29. CLASS NUMBER (0,n) <assigned to> BIBLIOGRAPHIC UNIT (0,n)
30. CONCEPT (0,n) <is related to> CONCEPT (0,n)
    relationship type (ex.: superconcept/subconcept, whole/part, type/token)
31. CONCEPT (0,n) <is part of> CONCEPT CLUSTER (1,n)
32. CLASS NUMBER RANGE (0,n) <is broader than> CLASS NUMBER RANGE (0,n)
33. CLASS NUMBER (0,1) <is within> CLASS NUMBER RANGE (0,n)
34. CLASS NUMBER (1,1) <authorized by> ORGANIZATIONAL TOOL (0,n)
    history note
35. BIBLIOGRAPHIC SET (0,n) <available from> BODY (0,n)
    (amount, currency)
    date range
36. BODY (0,n) <acquires> BIBLIOGRAPHIC COPY (0,n) <from> BODY (0,n)
    method of acquisition
    date of acquisition
    (currency, amount)
37. BIBLIOGRAPHIC COPY (1,n) <owned by> BODY (0,n)
    (start date, end date)
38. BIBLIOGRAPHIC COPY (1,1) <held in> COLLECTION (1,n)
39. SERIAL (0,n) <held in> COLLECTION (0,n)
    summary holdings statement
40. BIBLIOGRAPHIC COPY (0,n) <transferred to> BODY (0,n)
    transfer type
    transfer date
    due date
    returned date
41. BODY (0,n) <places hold on> BIBLIOGRAPHIC SET (0,n) <in> COLLECTION (0,n)
    hold-placed date
    hold-released date

A Converging Vision of Cataloging in the Electronic World

Ling Hwey Jeng

The concept of the library of the future has been discussed extensively throughout the past decade. The development of Internet resources currently underway is a step in this direction. However, methods of organizing electronic resources are still far from ideal. This paper will trace library cataloging from manual to electronic organization as well as review current attempts to organize electronic information objects and related issues. A vision of what cataloging could evolve into in the electronic world is proposed at the end, along with a discussion of what needs to be done to realize this vision.

In her recent review of libraries of the future, Drabenstott[1] collected twelve different definitions of the library of the future by scholars in the field, many of which shared common characteristics. This future library is one that emphasizes user access to information rather than ownership of materials. Instant information delivery through a system's direct links to remote files is assumed. The future library uses the maximum power of available electronic technology, constantly updates its records, and offers its users intelligent services. In this future library, access is not constrained by time or place. The future library's collection reflects true integration of traditional materials and electronic resources. A good definition of the library of the future is that it is a system that utilizes the maximum power of technology to link cataloging, indexing, and textual databases in an online network to provide users universal access to locally owned and remotely available information sources.

The key to this concept of the future library is electronic information. In an electronic world, organization occurs in whatever way is needed to provide quick and thorough access to information objects, regardless of physical constraint. One way to better understand the nature of organization in the future library is to understand what has been known as library cataloging. This is discussed in the following paragraphs. The paper then covers current attempts to organize electronic information objects. This is followed by some issues related to current practice in organizing electronic resources. A vision of what cataloging could evolve into in the electronic world is proposed at the end.

Characteristics of Traditional Cataloging

Ample literature exists on the description, principles, and practice of traditional cataloging. A cursory look at the definition of a catalog reveals some of the characteristics of traditional cataloging practice. The *Anglo-American Cataloguing Rules* defines a "catalog" as: "(1) A list of library materials contained in a collection, a library, or a group of libraries, arranged according to some definite plan. (2) A list of materials prepared for a particular purpose (e.g., an exhibition catalog, a sales catalog)."[2]

These two definitions show that traditional cataloging is based on physical formats, assumes homogeneous collections, and is primarily location bound. The emphasis in practice is on a uniform approach to organization and arrangement. Traditional catalogs facilitate known-item searching, category searching, and choice of entries, as spelled out by Cutter in 1876.[3] This is done through bibliographically significant access points such as authors, titles, and subject headings, in addition to some commonly used alphanumeric search keys, such as ISBN. An alternative is to use a subject approach to access within certain hierarchical structures, such as a classification system.

The indexing files for traditional catalogs are very predictable. Library bibliographic databases use inverted file structures with linear order in their arrangement.

Most of these characteristics, however suitable for the traditional world of bibliographic resources, do not apply as well to the organization of electronic information. As Erik Jul noted, "Books and similar things benefited from bibliographic description and access; Internet resources did not. Books and similar things could be located using the library catalog; Internet resources could not. . . . Two worlds arose: the library and its catalog and the Internet and its current, most favorite searching/browsing/grazing/burrowing scheme."[4] This by no means suggests that catalogers are an isolated group, unaware of the development of electronic technology and the changes in the library landscape. On the contrary, catalogers, especially those in academic libraries, were one of the first groups to embrace the Internet technology as part of their work routines. Through bibliographic utilities such as OCLC and RLIN, systems like the Internet, and organizations such as LC, catalogers have been actively engaged in e-mail exchanges related to their jobs and have participated in one of the largest electronic listservs, AUTOCAT. Cataloging records in MARC format have been made available via the Internet by several institutions, including the Australian Biblio-

Ling Hwey Jeng is Associate Professor, School of Library and Information Science, University of Kentucky; e-mail lhjeng00@ukcc.uky.edu.

graphic Network, the Library of Congress, and RLIN, which allow export of MARC records via FTP.

Catalogers have also taken advantage of Internet resources such as the World Wide Web to enhance their work. The Web allows them quick access to resources in other libraries. For example, the Web is a useful tool when foreign-language experts in other institutions need to be located quickly. The Web is also an excellent reference source when catalogers see the need to familiarize themselves with a new subject area. Finally, many cataloging tools are now updated in electronic format and made available on the Internet (e.g., LC Marvel, LC, OCLC FTP).[5]

Catalogers cannot ignore Internet resources without being deprived of opportunities for professional enhancement. From the beginning, catalogers have not realized the opportunities to shape the evolving electronic world or positioned themselves to lead in developing the structure and organization of Internet resources.

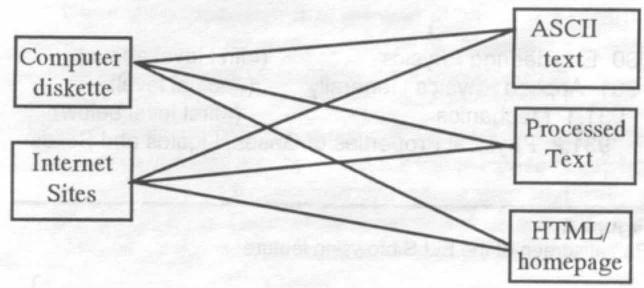
Current Status of Organization in the Electronic World

The electronic world is messy, volatile, and uncontrolled. For example, it is possible for multiple versions of an electronic information object to exist. The electronic text could be made available as a standalone ASCII, PostScript. or other file. The file can reside on a computer diskette or on a mini- or mainframe computer system as a file with a URL address and directory path. Many home pages not only contain text files but also serve as the front-end service for information systems (see figure 1).

So far, most attempts to organize Internet resources and provide users access have been done haphazardly. It is not difficult to find electronic indexes to Internet or other library resources created as independent projects by individuals with computer skills, creativity, and time to spare. Many such indexing systems are still at the development stage.

Monroe at North Carolina State University categorizes catalogs on or of the Internet into the following groups:[6]

1. *Catalogs of Internet resources*: These are catalogs created to organize only Internet resources, often with direct links to the sites of the electronic resources. Examples of this group include the Alex Catalog of Electronic Texts on the Internet[7] and the NEEDS courseware database of scientific and engineering courseware on the Internet.



Possible versions include:
archival ASCII files on computer diskette
archival processed files on computer diskette
archival ASCII files on FTP and gopher sites
archival processed files (PostScript, rtf) on FTP and gopher sites
WWW HTML archival files
WWW HTML service home pages

Figure 1
Multiple versions of an electronic information object

2. *Catalogs with vendor built-in Web interface but containing no Internet resources*: Catalogs in this group are mostly at the demonstration stage. Examples of vendors involved in such developments at the time of this study include SIRSI's WebCat, the DRA Library Web Information System, and VTLS's HTML Searching Interface.
3. *Catalogs with Web to Z39.50 gateway*: This group include library catalogs, such as that of the Library of Congress and Michigan State University's Zweb, that are equipped with Z39.50 client server software, which allows more user-friendly remote retrieval.
4. *Catalogs with Web interfaces for traditional materials*: This group of catalogs include that of Carnegie Mellon University and the NASA Lunar and Planetary Institute. They provide only surrogate information to traditional nonelectronic collection.
5. *Catalogs with gopher gateways*: This group includes the Alex Catalog of Electronic Texts on the Internet (gopher version), the Norwegian National Union Catalog, and the Eisenhower National Clearinghouse for Math and Science Education.

Internet Indexing Projects

Among the experiments, prototypes, and working projects, there are five worth noting: Yahoo at Stanford; Lycos at CMU; Nordic WAIS/World Wide Web Project;

930 Engineering Physics (=first level above)
931 Applied Physics Generally (=actual level)
931.1 Mechanics (=first level below)
931.2 Physical Properties of Gases, Liquids and Solids

Figure 2
Partial screen of the ELLS browsing feature

EELS, the Engineering Electronic Library at The Swedish University of Technology Libraries; and OCLC's NetFirst.

Yahoo indexes URL sites. Its owners claim that since humans do the indexing, quality control over selection of URL sites is maintained. The system indexes up to 100 URL addresses per person per day. The index, however, does not attempt to provide detailed information about the content of electronic files themselves.

Lycos uses software to automatically collect URL addresses for FTP, gopher, and news group sites. With the capability to index every and any digital document accessible, it provides the potential for quick access to the entire electronic world. However, it does nothing to help users determine the quality of the contents on the Internet.

The Nordic WAIS/World Wide Web Project is sponsored by NORDINFO and carried out by the National Technological Library of Denmark and Lund University Library. The objective is to explore the possibilities of improving navigation and searching on the Net. The results, according to its report, include integration of a library system into the Web; automatic detection and classification of WAIS databases, featuring a Web front end; automatic detection and indexing of Nordic Web pages; and support for multidatabase searching and relevance feedback.[8]

The EELS (Engineering Electronic Library) project, currently under development, is sponsored by the Swedish University of Technology Libraries.[9] The best-covered subject areas to date are physics, mathematics, energy, computer engineering, general engineering, and cold region research. One unique feature of this system is its classified approach to browsing EELS. The system provides the user with a hierarchical subject classificatory structure as a menu of choice, using Engineering Information Inc.'s EI classification (figure 2).

A new service announced in 1995 is OCLC's NetFirst. According to its news release, NetFirst automatically collects and verifies the Internet resources it gathers and utilizes conventional abstracting and indexing practices done by humans. It contains a database of Internet resources, including Web pages, interest groups, library catalogs, FTP sites, Internet services, gopher servers, and electronic journals and newsletters. Each record for an Internet resource contains the bibliographic citation, summary descriptions, and subject headings. Location information is included in the bibliographic records and can be used to connect users to resources of interest.[10]

These are intended to index electronic resources and to provide users quick access and information delivery. None of them involves any concerted efforts to practice library cataloging as it is currently understood. In all the projects, the emphasis is solely on fast indexing and speedy access. The issues most important to catalogers, namely quality and completeness of record description, are only peripheral, if addressed at all.

Cataloging Internet Resources

Electronic resources have been cataloged in libraries for more than a decade using *AACR2* chapter 9, but catalogers dealt with mostly static archival computer files. In 1992 the library cataloging community began to attempt to catalog electronic resources on the Internet. The following three projects are a significant departure from existing cataloging practice in that they expand the scope of traditional cataloging to include not only archival computer files but also dynamic electronic resources currently in use or under development. These three projects are the OCLC Internet Cataloging Resources Project, the University of Virginia Library Project, and the OCLC Cataloging Internet Project.

In their report, Dillon et al.[11] state that the objective of the OCLC Internet Cataloging Experiment is to study the nature of electronic information via remote access and its related problems. The project involved creation of MARC records for various types of Internet resources including source/system codes and text files. The records were then made available through remote access via FTP sites. The project resulted in a proposal for a new MARC field 856, approved by MARBI in 1995,[12] and the drafting of new cataloging guidelines for organizing Internet resources.

In 1992 the University of Virginia Library established an Electronic Text Center, which is charged with the task of digitizing the library's literature collection, including Old English texts, the *Chadwyck-Healy English Poetry Database*, the *Patrologia Latina* database, the second edition of the *Oxford English Dictionary*, and the complete works of Shakespeare and Thomas Jefferson. The digital text is marked up using SGML. As the document is marked up, a TEI header is created to record bibliographic data. Meanwhile, a MARC record is created as well for each electronic document using AACR2, LCRI, OCLC documentation. The result is a database of bibliographic records in MARC format.[13]

At the time OCLC finished its Internet Cataloging Experiment in 1993, a follow-up project, called the OCLC Cataloging Internet Project, was initiated. The project's goals included the identification, selection, and cataloging of Internet resources.[14] To date, the project has more than 170 participant libraries, each producing an average of two to three records per week. The resulting cataloging records comply with MARC and AACR2 standards and are suitable for loading into national, regional, or local library catalogs.

From examination of the MARC records, it is clear that the main types of the electronic resources cataloged under this project are Internet systems and services, and single bibliographic entities with URL addresses. A commonly recognized policy is to catalog those Internet resources that are locally owned, such as the home pages of the local library catalog, or those of local interest. Reference tools considered useful to local users are next on the priority list, as are professional tools available on the Internet. Participating libraries are also encouraged to create multiple 856 fields for multiple versions of an electronic.[15] Cataloging under the OCLC Internet Cataloging Project is done in accordance with *AACR2*, chapter 9 and their related *LCRIs*, MARC formats, OCLC's *Cataloging Internet Resources: A Manual and Practical Guide*,[16] and *Guidelines for the Use of Field 856* published by the Library of Congress.[17]

Current Issues

Compared to the various indexing projects described earlier, these three projects for cataloging electronic resources deal with only a small number of Internet resources. Each of the three projects produces a separate bibliographic database of MARC records, often without actual links to the electronic resources themselves. The lack of electronic link between records and their resources brings out some of the issues in current attempts to organize the electronic world.

Issue 1: The separation of the bibliographic record and its primary electronic source. Electronic files can be created by almost anyone who has a word processor or scanner, an Internet address, or a home page. With almost all current projects, the record created is separated from the electronic text itself, making it essential to establish a link between the primary text and the surrogate for each case.

Issue 2: The time and labor to create MARC cataloging records. There is no assurance that the cost of creating a record is worth doing so when the contents of electronic resources are in many cases unstable or even questionable.

Issue 3: The uncontrolled quality of Internet indexes. Most Internet indexes were created for the purpose of capturing the most addresses and providing the most links to electronic data. There is no consistency in indexing, nor is there a minimum standard for record description.

Issue 4: Duplicate effort in creating MARC records and SGML coding. Some systems (such as the University of Virginia Project) have tried to create both the SGML/TEI header and the MARC record for each electronic document; both contain bibliographic data. This represents duplication of coding and processing effort. An example is the treatment of language text. There are provisions in both SGML coding and MARC tags for language treatment; either is sufficient for the record. To have both created using separate standards is redundant.

Issue 5: The volatility of electronic resources. Because of the ease and speed of production of electronic resources, the index links created for accessing the resources often are just as volatile. No matter how quickly a system indexes Internet resources, the resources themselves appear and disappear even faster. As operating systems are upgraded and files moved to new sites and assigned new addresses, links between an electronic file and its bibliographic record can become invalid at any time without notice.

To solve this problem, some suggest letting users report the invalid links found. Others see the need to ask electronic publishers to notify libraries when changes in URL addresses occur. A third approach suggested by Dodge, Marx, and Pfeiffenberger is to use software to automatically check links regularly.[18] More recently, OCLC has announced the implementation of its Persistent Uniform Resource Locator (PURL) as an intermediate service to resolve frequent changes in URL links. PURL appears to be similar to a URN standard currently under development by Internet Engineering Task Force (IETF).[19]

Electronic Evolution

Given the issues, it is obvious that cataloging in the electronic world can no longer be done the same as it has been in the past. We must ask ourselves what we want to see in the future of cataloging and what the acceptable norm of organization for the electronic world is if not cataloging as usual.

One can begin answering this question by identifying the electronic entities involved in the process of organization. There are at least six, including local documents, remote documents, bibliographic data of any document, the citations or references used in a document, locally created surrogate files, and remotely accessible surrogate files.

With local or remotely accessible electronic documents, the document may be full text with possible illustrations and/or graphic elements. It is likely the documents will contain labels of some markup language, such as HTML. It is necessary to consider both bibliographic data already existing in the document as well as those that could be added during the cataloging process. When creating links, both the bibliographic records of the document and the bibliographic elements provided in the citations of a document can be used as indexing pointers to other related documents. The local surrogate files must be linked with the remotely accessible surrogate files. Without simultaneous access to both local and remote surrogate files, the system loses one of its most valuable functions as an electronic surrogate system.

In addition to the essential entities of electronic resources, one must also consider the following factors in the future of organizing electronic resources.

Creation of electronic files is a dynamic process. In the Web environment, a file is normally created on a local computer using a word processor. The file, once completed, is marked up using an HTML editor. Graphics and TEI headers are often inserted during the markup process. The file is then uploaded to a home page and given a URL address. Although many HTML files remain text-based files, many others contain additional hypertext links embedded in the text including the citations that serve as pointers to local and remote URL addresses of other related files.

One current feature of electronic indexes that must remain in a successful future cataloging system is fast indexing and access. With the accelerating rate in which electronic files and home pages are created on the Internet, the current level of productivity for creating MARC records for electronic resources is unacceptably low. In increasing productivity, however, the quality of the records and the indexing structures must be maintained. Standards should be enforced to ensure that a minimum level of bibliographic data is provided in each description. Sophisticated search keys through controlled vocabularies, classification systems, and name authority control must be made available to the users through an intelligent interface.

Defining the Electronic Unit of Control

Another critical issue in dealing with electronic information is the question of the unit of control. Traditional cataloging prescribes that cataloging can only be done with an item in hand. The physical format is the basis for determining what cataloging rules to apply and what kind of record to create for the item.[20] This fundamental principal became irrelevant for organizing electronic information, as seen in discussions that resulted from reviews of a bibliographic record created for the AUTOCAT listserv on the OCLC Internet Cataloging Project.[21] It is evident in recent discussions of the fluid, dynamic nature of most electronic information objects[22] that such objects must be defined by the way they are processed and stored, which is as computer files, whether on a diskette or at a URL site. A computer file with a unique file ID (or file name) and location information is a cohesive unit that can be created, stored, retrieved, processed, and transferred. Although in many cases a collection of files with a common link (through usage or content relevancy) must be treated together in the organizing process, each individual file must be maintained separately. The issue of defining electronic information objects as files will be further discussed in the author's other publications. Suffice it to say that the file as the unit of control must be the basis for considering organizing methods in the electronic world.

A Converging Vision

Based on all the factors discussed, a converging vision is proposed here as the direction for cataloging in the electronic world.

As an electronic document is prepared to be added to the electronic world, such as an Internet URL site, it is tagged in an electronic markup language such as HTML. In place of the TEI header commonly used in today's markup process, an expanded document markup header can be created with human intervention as a cataloging record of the document. The human markup editor, armed with cataloging knowledge and familiarity with markup language syntax, can provide necessary information for bibliographic description. The electronic system automatically verifies authoritative forms of access points using its intelligent knowledge base. Meanwhile, hypertext links are created for the electronic text by marking keywords within the text as well as its citations. At this point, the process of creating the electronic information object is complete.

The next step is to store the electronic file at a URL address with other local files. During this process, the system extracts the catalog record from the file's header, which is automatically added to the system's local catalog. The local catalog is linked with other local and remote surrogate files through home pages, which allows users to get instant access to both. With the connection completed, the catalog essentially becomes an invisible network with links between local library resources and remote databases, and allows users to access information without regard to its physical or temporal constraints.

Whenever the file is in need of revision, the markup language editor not only updates the content of the file but also updates the expanded cataloging header at the same time. When the document is moved, its URL address is changed, either automatically or by the (human) markup editor. When the file is deleted, the expanded cataloging header disappears with the file from the system, thus closing the link between the file and the rest of the electronic world with a message that the file is no longer available.

Characteristics of the Vision

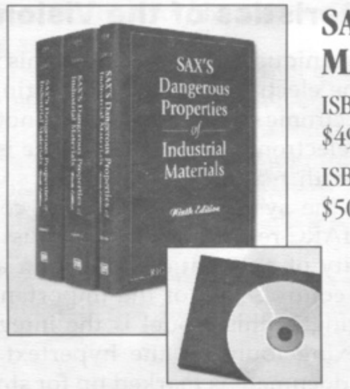
There are several unique characteristics of this vision for cataloging for the electronic world. Cataloging records exist with(in) electronic source files and are not separate from them. The electronic catalog in essence is a collection of records with real-time links to the source files themselves. For the system to generate a correct TEI header for the MARC record, the system must also possess the capability of automatic verification and automatic authority control. One of the important components of cataloging in this model is the integration of associative indexing found in the hypertext environment. When the document is marked up for storage and presentation in electronic form, the text, as well as its citations, is also marked up with hypertext nodes to be used as indexing pointers. In this vision of future cataloging, cataloging and indexing are not two separate steps, but one integral process done at the time of markup editing; the markup language editor is the cataloger and the indexer. In effect, the three functions become one. This process of markup/cataloging/indexing can be decentralized, done by individuals in or outside of an institution such as a cataloging department.

What Does It Take to Get There?

Catalogers have been using electronic resources as the prime avenue for updating their professional knowledge for some time. Many cataloging tools are regularly updated and posted on a Web site for free access, and more will be so in the future. Catalogers have also utilized electronic resources as a tool for shared cataloging and will continue to take full advantage of these features.

The key to realizing this vision of the future organization of electronic resources is to integrate the tasks of cataloging, indexing, and markup editing. In order to keep up with the growth of electronic resources, greater flexibility in traditional cataloging quality and standards must be implemented to improve the productivity. This flexibility might seem to compromise current cataloging quality, but the benefit of such compromise will be more records and speedier access. Associative indexing will become an integral part of cataloging knowledge. Catalogers not only have to be comfortable with using subject headings lists or classification systems, they also will have to be comfortable with alternative indexing structures, such as hypertext indexing and electronic markup languages. The ability to use markup languages, such as SGML or HTML, is essential for catalogers to participate in the production process of

Distinct, Comprehensive, Powerful... that's what sets VNR's chemical and environmental references above the rest (Available in print and CD-ROM)

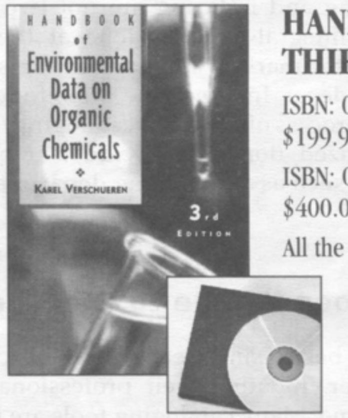


SAX'S DANGEROUS PROPERTIES OF INDUSTRIAL MATERIALS, NINTH EDITION by Richard J. Lewis, Sr.

ISBN: 0-442-02025-2 (Three-Volume Set) **Available now**
$499.95 ©1996

ISBN: 0-442-02335-9 (CD-ROM) **Available now**
$500.00 ©1996

A vast, reliable source of critical information on health and safety data, regulatory standards, toxicity, carcinogenicity, and physical properties of over 20,000 chemical substances. New: • 2,000 completely new chemicals, drugs, and biological agents • 14,000 updated entries • New or additional physical data added for 9,000 entries • CAS Registry Numbers for more than 18,000 entries • 100,000 domestic and international synonyms • 3 Instant Locator Indexes • Streamlined toxicity data • OSHA, ACGIH, MAK exposure limits • NIOSH REL workplace air limits • New advanced safety profiles for handling spills, accidental exposures, and fires • Carcinogens and reproductive effects • New European Data. Also available in CD-ROM.

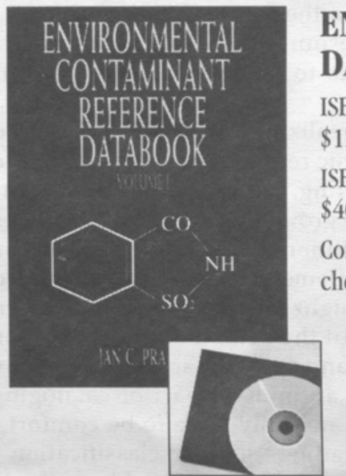


HANDBOOK OF ENVIRONMENTAL DATA ON ORGANIC CHEMICALS, THIRD EDITION by Karel Verschueren

ISBN: 0-442-01916-5 (Print Version) **Available April 1996**
$199.95 ©1996

ISBN: 0-442-02421-5 (CD-ROM) **Available June 1996**
$400.00 ©1996

All the information you need to prudently use potentially dangerous chemicals is in this fully revised edition of our classic handbook. Pesticides, detergents, phthalates, polynuclear aromatics, and polychlorinated biphenyls are all investigated in detail. Includes physical/chemical properties; air, soil, and water pollution factors; aquatic toxicity, and biological effects; odor thresholds; sampling and analysis data; and structural formulas of over 3,000 chemicals. Tables have been refined to focus on environmentally related materials. Also available in CD-ROM.

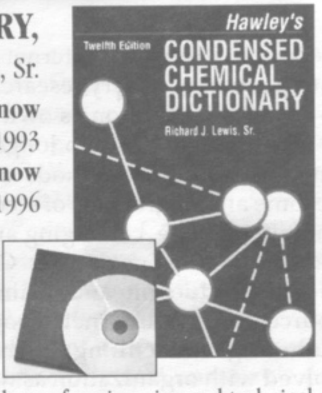


ENVIRONMENTAL CONTAMINANT REFERENCE DATABOOK, VOLUMES I AND II by Jan C. Prager

ISBN: 0-442-01918-1 (Vol. I), ISBN: 0-442-01969-6 (Vol. II) **Available now**
$129.95 (each volume) ©1995, 1996

ISBN: 0-442-02420-7 (CD-ROM) **Available now**
$400.00 ©1996

Comprehensive coverage of environmental effects of regulated chemicals. Identifies regulated chemicals and chemicals of special interest to environmental and safety professionals due to widespread contamination or considerable hazard. A reference compendium of physical, chemical, and biological dangers of environmental chemicals and selected simple and complex mixtures. Listings include CAS number; SAX number; sampling, analysis, and detection limits; environmental transport, fate and effects; structural formula; molecular formula; common uses; manufacturers; odor thresholds; regulatory jurisdictions and authorities; standards; reactions; monitoring methods; international data; synonym index. Also available in CD-ROM.

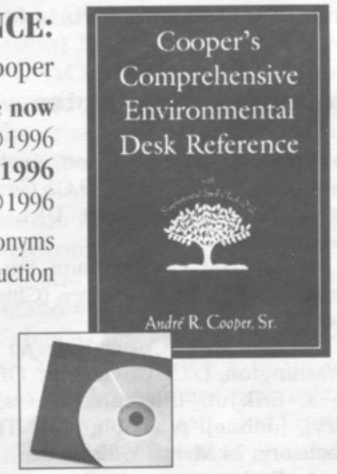


HAWLEY'S CONDENSED CHEMICAL DICTIONARY, TWELFTH EDITION by Richard J. Lewis, Sr.

ISBN: 0-442-01131-8 (Print Version) **Available now**
$79.95 ©1993

ISBN: 0-442-02419-3 (CD-ROM) **Available now**
$300.00 ©1996

The most widely recognized dictionary of industrial chemicals, terms, processes, reactions, and related terminology, Hawley's is an essential reference for all levels of staff in industries where chemicals are used — from management and administrative people to marketing and sales to technical and scientific staff. Hawley's presents three distinct types of information: (1) descriptions of chemicals, raw materials, manufacturing processes, and equipment; (2) expanded definitions of chemical entities, phenomena, and terminology used in every phase of engineering and technical development; and (3) descriptions or identifications of a wide range of trademarked products. Also available in CD-ROM.

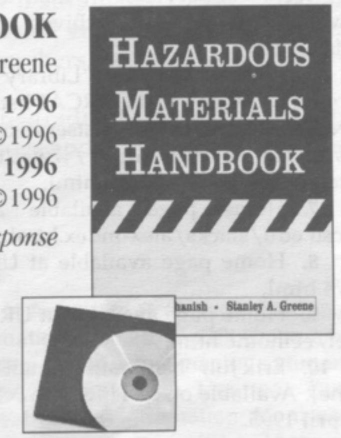


COOPER'S COMPREHENSIVE ENVIRONMENTAL DESK REFERENCE: WITH SUPPLEMENTAL SPELL CHECK DISK by André R. Cooper

ISBN: 0-442-02159-3 (Print Version) **Available now**
$99.95 ©1996

ISBN: 0-442-02161-5 (CD-ROM) **Available August 1996**
$200.00 ©1996

A unique comprehensive reference with more than 1,000 pages of key terms, over 5,000 acronyms and useful, detailed information on: • Superfund • Environmental engineering • Biology • Construction • Wetlands • OSHA • Threatened and endangered flora and fauna • and International organizations and treaties. Information is supplemented by maps, charts, tables, and diagrams, as well as crucial agency contact information to give the professional access to environmental officials who can offer advice and assistance. Destined to become a classic work, essential for all environmental, hazardous waste and safety engineers and managers, industrial hygienists, and environmental lawyers, as well as for students in these fields. Also available in CD-ROM.



HAZARDOUS MATERIALS HANDBOOK by Richard P. Pohanish and Stanley A. Greene

ISBN: 0-442-02212-3 (Print Version) **Available July 1996**
$149.95 ©1996

ISBN: 0-442-02422-3 (CD-ROM) **Available September 1996**
$400.00 ©1996

This corrected, updated and expanded version of the U.S. Coast Guard's *Chemical Hazards Response Information System (CHRIS) Manual* features widely used and transported industrial materials including, practical hands-on data and technical data as well as chemical properties. This book is a portable and easier to use version of the *CHRIS Manual.* This new resource covers more than 1,240 substances, while also improving on the original by including a CAS index, synonym index, and a helpful identification section with new fields and vital ID information moved to the front of each record for easy accessibility. Also available in CD-ROM.

Order from your local wholesaler or order direct from:
Donald O'Connor, National Accounts Manager, Van Nostrand Reinhold
212-780-6129 (Phone) • 212-254-9499 (Fax)
doconnor@vnr.com (E:mail)

Van Nostrand Reinhold
115 Fifth Avenue, New York, NY 10003

VNR

electronic resources. Internet searching will become an integral part of library research, as important as learning to search OCLC or one's own online catalog.

Cataloging can no longer be a rigid process separated from other methods of organization, but must become an integral part of electronic authorship/editorship. In practice, cataloging and markup editing should be merged into one process. Catalogers will be involved in the production and maintenance of electronic resources on the Internet. People currently involved in authoring and editing Internet resources will be involved with organization as well. The converging vision of organization in the electronic world represents an integrated approach to production and organization of electronic resources that moves catalogers from the sidelines of the electronic world to the center stage.

References and Notes

1. Karen M. Drabenstott, *Analytical Review of the Library of the Future* (Washington, D.C.: Council on Library Resources, 1994). Also available at URL ftp://sils.umich.edu/pub/papers /CLR.

2. Michael Gorman and Paul Winkler, *Anglo-American Cataloguing Rules*, 2d ed. rev. (Chicago: American Library Association, 1988), 616.

3. Charles A. Cutter, *Rules for a Dictionary Catalog*, 4th ed. (Washington, D.C.: Govt. Print. Off., 1904).

4. Erik Jul, "Integrated Access," [e-mail to INTERCAT listserv], [online]. Available on INTERCAT archives at listserv @oclc.org, 24 March 1995.

5. Barbara Stewart, "Summary of Replies to WWW Access for Technical Services," [e-mail to LITA-L listserv], [online]. Available on LITA-L archives at listserv@uicvm.bitnet, 27 March 1995.

6. Hunter Monroe, "Library Catalogs of Internet Resources," [e-mail to INTERCAT listserv], [online]. Available on INTERCAT archives at listserv@oclc.org, 18 April 1995. Also available at URL http://www.lib.ncsu.edu/staff/morgan/alcuin/wwwed-catalogs.html.

7. Home page available at URL http://www.lib.ncsu.edu/stacks/alex-index.html.

8. Home page available at URL http://www.ub2.lu.se/W4.html.

9. Home page available at URL http://WWW.UB2.lu.se/eel/eelhome.html.

10. Erik Jul, "NetFirst," [e-mail to INTERCAT listserv], [online]. Available on INTERCAT archives at listserv@oclc.org, 7 April 1995.

11. Martin Dillon and others, *Accessing Information on the Internet: Toward Providing Library Services for Computer-Mediated Communication* (Dublin, Ohio: OCLC Online Computer Library Center, 1993). Available at URL ftp://ftp.rsch.oclc.org/pub/internet_resources_project/report.

12. Library of Congress, Network Development and MARC Standards Office, *Guidelines for the Use of Field 856* (Washington, D.C.: Library of Congress, 1995). Also available at URL http://lcweb.loc.gov/marc/856.guide.html.

13. Edward Gaynor, "Cataloging Electronic Texts: The University of Virginia Library Experience," *Library Resources & Technical Services* 38 (Dec. 1994): 403–13.

14. Erik Jul, "Getting a Sense of Involvement," [e-mail to INTERCAT listserv], [online]. Available on INTERCAT archives at listserv@oclc.org, 13 April 1995.

15. Vanna Sha, "Guidelines for Cataloging Electronic Resources," [e-mail to INTERCAT listserv], [online]. Available on INTERCAT archives at listserv@oclc.org, 28 March 1995.

16. Nancy B. Olson, *Cataloging Internet Resources: A Manual and Practical Guide* (Dublin, Ohio: OCLC Online Computer Library Center, 1995). Also available at URL http://www.oclc.org/oclc/man/9256cat/toc/html.

17. Library of Congress, Network Development and MARC Standards Office, *Guidelines for the Use of Field 856.*

18. C. Dodge, B. Marx, and H. Pfeiffenberger, "Web Cataloguing Through Cache Exploitation and Steps Toward Consistency Maintenance," *Computer Networks and ISDN Systems* 27 (1995): 1003–8. A WWW link checker, called WebWatcher, is also available at URL http://www.specter.com/users/janos/specter/webwatcher.html.

19. See for example, Erik Jul, "PURLs" [e-mail to INTERCAT listserv], [online] (available on INTERCAT archives at listserv@oclc.org, 5 January 1996); and Kenneth E. Rowe and Charles K. Nicholas. "Reliability of WWW Name Servers," *Computer Networks and ISDN Systems* 27 (1995): 773–80.

20. Gorman & Winkler, *Anglo-American Cataloguing Rules.*

21. See for example, Charles F. Bearden, "Re: Cataloging Record for Autocat listserv" [e-mail to INTERCAT listserv], [online] (available on INTERCAT archives at listserv@oclc.org, 27 July 1995); Marty Rosen, "Re: Cataloging Record for Autocat listserv" [e-mail to INTERCAT listserv], [online] and Lorraine Moore, "Re: Cataloging Record for Autocat listserv" [e-mail to INTERCAT listserv], [online]. Both available on INTERCAT archives at listserv@oclc.org, 1 August 1995.

22. For a discussion of the definitions of information objects, see Ling Hwey Jeng, "From Cataloging to Organization of Information: A Paradigm for the Core Curriculum," *Journal of Education for Library and Information Science* 34 (Spring 1993): 113–26. Other recent discussions on this issue include Linda Schamber, "What Is an Electronic Document?" A contributed paper presentation at 1995 ASIS Mid-Year Meeting, Minneapolis. May 24–26, 1995.

PACE: A Browsable Graphical Interface

Jamshid Beheshti,
Valerie Large, and Mary Bialek

PACE (Public Access Catalogue Extension) is an alternative interface designed to enhance online catalogs. PACE simulates images of books and library shelves to help users browse through the catalog. PACE was tested in a college library against a text-based, online public access catalog (Best-Seller) in a real operational environment. The results show that a simple browsable retrieval interface performed as well as a second-generation OPAC in terms of retrieval speed and search success. The overwhelming majority of students, however, preferred the browsing capability of PACE through the familiar metaphor of books and library shelves to a text-based OPAC.

Today many information resources can be accessed through the virtual library, including many online public access catalogs (OPACs) with sophisticated retrieval engines. Research shows, however, that novice users encounter difficulties in interacting with these systems (Borgman 1986). Yee (1991) has summarized these problems as finding appropriate subject terms, large numbers of hits and failure to reduce the retrieval sets, zero hits and failure to increase the retrieval sets, failure to understand cataloging rules, and spelling and typographical errors. In addition, lack of understanding of indexes, files, and basic database structure leads to the use of articles, stop words, entering author's first name before last name, and hyphenation problems. The interface and retrieval systems can also be a source of potential problems. Complex interfaces and the need for training and relearning when used infrequently, incomprehensible error and help messages, problems associated with displaying records and difficulties with Boolean logic have compounded the obstacles encountered by novice users. These restraints have prompted one researcher to state that the second-generation OPACs are "powerful and efficient but are dumb, passive systems which require resourceful, active, intelligent human searchers to produce acceptable results" (Hildreth 1989).

Suggested Solutions

To address the needs of end users and to alleviate the mentioned difficulties, many researchers have conducted experiments to enhance and improve OPACs. In general, research in this area may be divided into two broad categories: enhancing MARC records, and improving the retrieval engine and the interface. Many researchers (Cochrane 1986, Drabenstott et al. 1990, Chan 1990, Pejtersen 1989, Lawrence 1985) have demonstrated the value of augmenting MARC records and the online catalog with various schemes such as the inclusion of the Dewey Decimal Classification (DDC) schedules and relative indexes. Projects in Carnegie Mellon University and other institutions have attempted to enhance bibliographic records by adding table of contents of books (Greenwood 1989, Michalak 1990, Posey and Erdmann 1986). The addition of new information to MARC, however, can be very costly and unsuitable for individual libraries.

Advancing the retrieval engine has involved ranking of documents, weighting index terms and automatic spell checking. Projects such as OKAPI (Online Keyword Access to Public Information) have attempted to use these techniques to build OPACs which do not require any user training (Greenwood 1989) Other researchers have strived to enhance the retrieval system by adding visual interfaces which use icons, graphical user interfaces, hypertext links and multimedia to help end users. Client/server architectures are used to take advantage of the power of today's microcomputers to present users with alternative interfaces. Kid's Catalog (Busey and Doerr 1993), OASIS (Buckland et al. 1992), Multimedia Visualizer (Lee 1991), the Science Library Catalog Project (Borgman et al. 1995) and XOkapi (Hancock-Beaulieu et.al. 1995) are just a few examples of these new interfaces. These OPACs are utilizing concepts that move users closer to direct manipulation of objects or documents. Hildreth (1989) states that these OPACs are more intuitive as the objects are manipulated directly "avoiding previous layers of mental encoding/decoding and indirect representation searchers are usually required to pass through."

Although OPACs are gradually incorporating object manipulation techniques and browsing capabilities,

Jamshid Beheshti, Ph.D., is Associate Professor, Graduate School of Library and Information Studies, McGill University. At the time of this study, **Valerie Large** was a master's student in Computer Science, Concordia University, Montreal. Currently, she is Software Analyst at DRA Information, Montreal. **Mary Bialek** is Computer Services/ILL Head (Acting), Humanities and Social Services Library, McGill University, Montreal.

Correspondence should be addressed to Jamshid Beheshti, Associate Professor, Graduate School of Library and Information Studies, McGill University, 3459-McTavish St., Montreal, Quebec, Canada H3A 1Y1; email injz@musicb.mcgill.ca.

one study indicates that library patrons *browse the actual shelves even after consulting the online catalog* (Markey 1984). The data from another study shows that between 30 percent and 45 percent of all searches which start from the online catalog, regardless of type of search, are concluded with browsing the library shelves (Hancock-Beaulieu et al. 1990).

Browsing

Browsing is an integral part of human behavior. Many life necessities are obtained by "browsing enormous, ever-changing ranges of goods in a multiplicity of locations" (Poulter 1993). It can occur in individual books, at library shelves, or in catalogs. Marchionini (1995) states that browsing is a "natural and effective approach to many types of information-seeking problems." Browsing has gained much attention recently; Chang and Rice (1993) have reviewed over one hundred publications to arrive at a conceptual framework and a taxonomy for browsing, and Marchionini (1995) has devoted one-third of his recent book to this subject in an attempt to outline design criteria for browsing in online systems.

Many OPACs offer both querying and browsing. While querying involves exact keyword or phrase matching by utilizing Boolean and proximity operators and other features of an online system, browsing consists of mainly scanning lists. Querying requires a large amount of cognitive load, compelling users to devise search strategies, conceptualize appropriate search terms, cope with Boolean logic, and manage large or zero number of hits. As more features are added to online systems to increase their retrieval effectiveness, the cognitive load on the user rises proportionally (Hancock-Beaulieu et al. 1995). Browsing, however, requires a "smaller cognitive load than analytical search strategies do" (Marchionini 1995, 103). It is easier to recognize information presented in a familiar metaphor, such as browsable lists, than to recall it from memory (Borgman et al. 1995). In many situations, novice clients may use a combination of both search techniques (Belkin et al. 1993).

In an online environment, browsing is primarily *visual* and *is dependent on patterns and shapes* presented on the screen (Cove and Walsh 1988). Direct manipulation is more important in browsing than other information-seeking strategies "because it involves cognitive, perceptual, and motor systems concurrently and continually" (Marchionini 1995).

One method of visualizing information is by simulating physical entities such as books, shelves, and physical layout of the library. The advantage of direct simulation is that users are presented with familiar mental images with which they have had contact since childhood. This recognition of familiar environment allows clients to "concentrate more cognitive resources" on the information-seeking activity (Marchionini 1995). Bates (1989) states that "creating a virtual physical layout on the screen may make it easier for the searcher to think of moving among familiar categories of resources in an information retrieval system, in the same manner in which they move among resources in the actual library."

Public Access Catalogue Extension

Paul Saffo, a research fellow at the Institute for the Future in Menlo Park, California, suggests that we are in a state "between two revolutions—one of print, not quite spent, and another of electronics, not quite underway" (Wahlde and Schiller 1993). Online catalogs also need this transitional state. In an attempt to bridge the gap between the electronic information and "a heritage of hundreds of years dealing with print" (Li 1991, 262), a new browsable graphical interface has been designed. Public Access Catalogue Extension (PACE) presents MARC records in a simulated environment, mimicking users' mental images of books and libraries, in an attempt to reduce their cognitive burdens (Beheshti 1992). PACE is designed to augment or extend the capabilities of existing online systems, not to replace them. Users should be able to choose between a traditional OPAC interface or PACE or a combination of the two.

Simulation data are derived from the physical description of books, including number of pages and height embedded in MARC records. This information is used to generate an image of a book, the dimensions of which are proportional to the actual size. The spine of each book appears on the screen as a three-dimensional image with call number and full or truncated title. Each book is assigned a color using a hashing algorithm based on the publisher and title in order to distinguish it from adjoining books. About ten books are displayed on one screen on a simulated shelf. Depending on their width (number of pages), fewer or more books may appear on the screen. A mouse is used for navigation, allowing the user to click on the spine of a book so that a simulated "title" page is displayed. This page contains the full title, author, call number, publisher, publication date, and LC subject headings derived from MARC record.

In the first phase of the PACE project, the basic interface without a retrieval engine was designed. Two different versions were created for testing the effectiveness of the interface. The first version (PACE1) used multishelf and single-shelf displays. It did not include

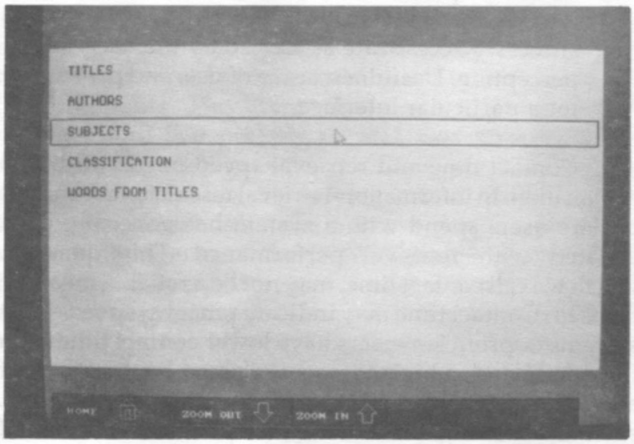


Figure 1
PACE List Options Screen

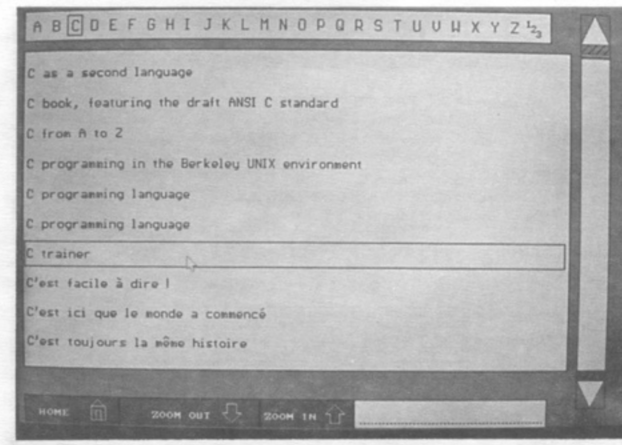


Figure 2
PACE Titles Screen

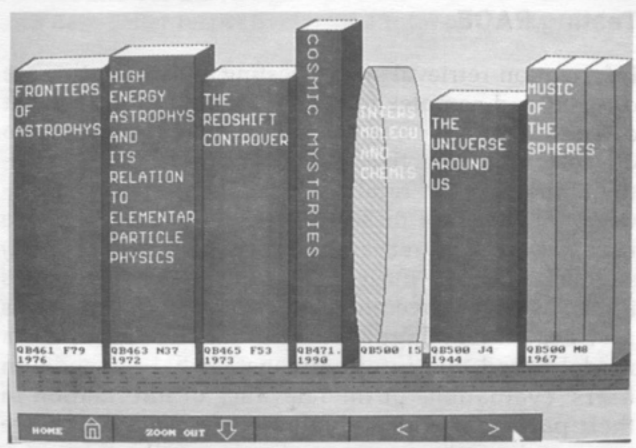


Figure 3
PACE Bookshelf Screen

any menus or navigational guides. The second version (PACE2) provided users with both a menu and a shelf display. The basic difference between the two versions was in their approach in retrieving information; PACE1 relied on a purely sequential approach where users had to move along the shelves in a linear fashion. PACE2 was a hybrid system providing users with random access points to the shelves. Both interfaces were tested against a second-generation OPAC (NOTIS). The results of a limited small-scale experiment indicated that users preferred the visual interface for accessing information to searching a command-driven text-based OPAC. The results also showed the need for a navigational tool or retrieval engine (Beheshti et al. 1994).

This is a report on the results of the second phase of the project. During this phase, a browsable retrieval engine was designed and added to the interface to produce a functional system. Since previous studies indicate that typing, keyboarding, and spelling may pose some problems for novice users (Solomon 1993, Connaway et al. 1995), PACE utilizes only a mouse for navigation without any input from the keyboard. Five indexes are created to facilitate retrieval: *Titles, Authors, Subjects, Classification,* and *Words from Titles.* The first three indexes or lists, as they are referred to in the interface, are based on the information from MARC fields. *Classification* is the general Library of Congress Classification scheme. *Words from Titles* is actually a KeyWord In Context index (KWIC) generated from titles (figure 1).

Each list is presented alphabetically for users to scroll through. To increase the speed of browsing, letters of the alphabet appear at the top of the screen indicating the first letter of a particular index (figure 2). While other schemes for increasing browsing speed have been suggested (see Allen 1993), it was decided that a simple approach such as the use of the alphabetical division would be sufficient for the purposes of this study. A vertical scroll bar is then utilized to move up and down the list. All the indexes or lists have the same interface to maintain consistency (Shneiderman 1992). Once a particular item has been chosen from any one of the five indexes, the bookshelf appears on the screen with an arrow indicating the selected book (figure 3). By clicking the left mouse button on the spine of any book on the shelf, a "title" page can be viewed (figure 4).

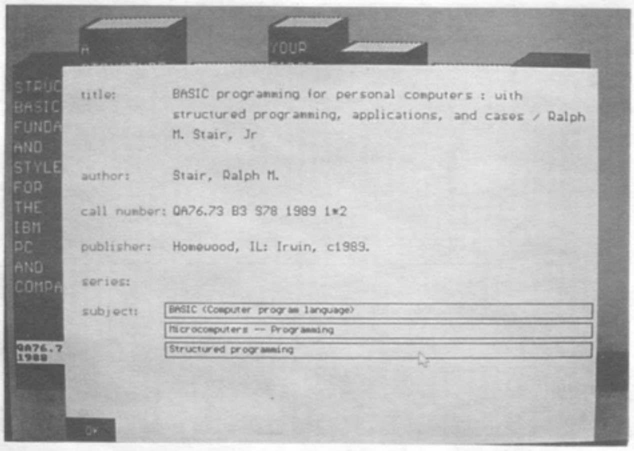


Figure 4
PACE "Title-page" Screen

Testing PACE

Information-retrieval system testing and evaluation are difficult and complex tasks, particularly in the context of a new interface. Recently attempts have been made to develop more effective methodologies than the traditional measures of recall and precision for testing systems. The focus of these new methodologies is on the user as well as the retrieval mechanism. Meadow (1992) suggests that the "most useful of the current measures if the objective is overall assessment of an IR process" is user evaluation. Tague and Schultz (1989) propose a model based on "informativeness" that depends on users' evaluations of the relevance of information in their particular context. Other researchers also place users at the center of any evaluation methodology for assessing online catalogs (Hancock-Beaulieu et al. 1990). Shneiderman's (1992) "five measurable human factors" rely on the user's evaluation of the system. These factors include success rates in searching for open-ended and closed questions, time to learn the new system, speed of performance, rate of errors by users, and satisfaction rates.

The main purpose for testing PACE is to determine its effectiveness as an alternative interface for searching online catalogs. Effectiveness has been defined in terms of factors relating to the interaction between users and OPACs. These factors, which have been mentioned in the above studies and others (see Chen and Dhar 1991) are:

1. contact time: The amount of time spent with the system;
2. retrieval speed: The amount of time needed to retrieve the first item of interest;
3. success: Success rate as judged by the user; and
4. perception: Usefulness, ease of use, and preference for a particular interface.

Contact time and retrieval speed need additional explanation. In information-retrieval research, the amount of time users spend with a system has generally been equated with retrieval performance. This quantity, which we call contact time, may not be a reliable measure by itself. Contact time may indicate retrieval speed—fast systems or proficient users have lower contact times. On the other hand, a high contact time may be the result of a user's decision to continue an activity on the system, such as browsing, and it may be a positive indicator of the system's usefulness. Retrieval speed as defined here, however, shows how fast the first item of interest may be retrieved. The assumption is that users start searching the catalog for a subject of interest. They locate one or more items, examine them and if useful, they proceed to browse the catalog for other relevant items. The faster they can find that first item of interest, the more time they have to browse for other relevant publications. Once the first source of information has been discovered, the decision to continue at the OPAC or at the shelf depends on the user.

Success rate has been defined in terms of items found during the search. Users may evaluate the results of a search as positive, i.e., finding the required information, negative, or undecided as to whether results are useful. Perception consists of three measures: usefulness of the system for a particular query, ease of use of the system, and the user's preference for an interface. Each factor is measured for two interfaces—PACE and a second-generation OPAC, Best-Seller.

Methodology

The second phase of PACE was tested under realistic operational conditions. Controlled experimentation can be more rigid and strict than operational testing; however, the artificial nature of the search topic hinders its usefulness. Under the experimental method, search queries may not represent real information needs. End users may not be able to judge reliably the relevance of search results to *their particular information needs*. Since the study of user behavior is a crucial aspect of this research, operational testing was preferred. Therefore the tests were conducted in "an operational environment, with real users with real, live problems; a live database . . ." and a functional system (Robertson and Hancock-Beaulieu 1992).

Environment
Testing was conducted at Vanier College, an English-language community college offering pre-university courses and career programs, with a total enrollment of 5,610 students. The library has a collection of 94,000 volumes and 500 periodicals and was recently automated by Best-Seller. OPAC terminals are available for students and staff in the reference area of the library. Best-Seller, which is based on SpeedWare fourth-generation language, has full retrieval capabilities including Boolean and proximity operators, keyword searching, and browsing features. Although it utilizes menus and function keys in a bilingual environment, it is primarily a command-driven system.

Database
The automated system at Vanier's library allows importing and exporting of MARC records. With the aid of the library and Best-Seller staff, the entire database consisting of approximately 92,000 records was transferred to a microcomputer. After initial clean-up, all the MARC records were imported into PACE.

Subjects
The majority of students entering Vanier have completed grade eleven in local high schools. Library staff provide bibliographic instructions through workshops and simple exercises handed out at the beginning of the school year in a required English course. According to the staff, most students have a very limited knowledge of libraries and are novice users of OPACs. For the purposes of this study, these students may be classified as *non-subject-specialized end users* (Meadow, Marchionini and Cherry 1994).

Procedure
In early September 1994, pretesting of the procedure and survey instrument was conducted. The actual testing lasted about two weeks. Two research assistants were situated in the reference area of the library where OPAC terminals are located. As students approached the online terminals to search the catalog, they were asked by one of the research assistants if they were willing to participate in the study. They were then assigned randomly to one of the two experimental groups; 95 searched PACE and 56 searched Best-Seller for their particular information needs. Since PACE is designed for browsing, assistants were instructed to choose those students who were not searching for exact known items such as materials on reserve or reading lists.

Each student was interviewed before and after the search by one of the two research assistants. A structured interview schedule was used to record students' previous online experience, library use habits, and demographic data. Students then proceeded to conduct their searches without any interference from assistants. Another structured interview schedule was used upon the completion of the search to solicit students' evaluation of and comments about the particular system they had used.

Although students in the Best-Seller group were familiar with the system, almost all were novice searchers. Students in the PACE group were provided with individual instructions, which on average lasted for three minutes. In general, the learning curve for PACE can be assumed to be negligible.

Since PACE is designed for browsing purposes, research assistants attempted to choose mainly those students who needed to conduct subject rather than known-item searching. Some of the data in the original samples, however, had to be eliminated from analysis. A number of students in both groups were actually conducting known-item searching, which research assistants were unable to verify at the time of interviews. The final counts for the two groups were 79 and 39 for PACE and Best-Seller respectively, for a total of 118 students.

Results

Testing for similarities among experimental groups
Since many variables can influence the outcome of retrieval testing, it is imperative that the two experimental groups do not significantly differ from each other. Gender, program of study or discipline, year of study, first language spoken, frequency of usage of computers, libraries, and OPACs were recorded for each student as independent variables (see Borgman 1989). Table 1 shows summary data for all variables collected at the time of interviews. Although there are some variations between the two groups, results of statistical analyses indicate that there are no significant differences in regard to gender ($\chi^2 = 0.5187$, df = 1, p = 0.4714), program of study ($\chi^2 = 0.8661$, df = 4, p = 0.9294), and first language spoken ($\chi^2 = 1.5402$, df = 2, p = 0.4630). Proportionally, there are more second-year students in the Best-Seller group than in the PACE group; however, this difference is not statistically significant ($\chi^2 = 1.9275$, df = 2, p = 0.3815). Similarly, the variations in frequency of computer usage ($\chi^2 = 2.1326$, df = 2, p = 0.3443), library usage ($\chi^2 = 0.7240$, df = 2, p = 0.6963), and OPAC usage ($\chi^2 = 3.5708$, df = 2, p = 0.1677) are not significant. The groups were also comparable in their topic knowledge domains prior to searches ($\chi^2 = 1.4076$, df = 2, p = 0.4947).

Contact Time
The total amount of time each student spent searching the catalog was recorded. Although times varied consid-

Table 1
Similarities between Experimental Groups

Variables	Categories	PACE No.	PACE %	Best-Seller No.	Best-Seller %
Gender	Female	39	49	22	56
	Male	40	51	17	44
Program of study	humanities	11	14	4	10
	social science	24	31	14	36
	natural science	23	30	11	28
	commerce	14	18	8	21
	computer science	6	8	2	5
First language	English	42	53	16	41
	French	13	17	8	21
	other	24	30	15	39
Year of study	first	30	39	11	29
	second	34	44	22	58
	third or more	13	17	5	13
Computer use	every day	20	25	5	13
	few times/week	28	35	16	43
	seldom	31	39	16	43
Library use	every day	13	17	5	18
	few times/week	41	52	12	43
	seldom	25	32	11	39
OPAC use	every day	6	8	2	5
	few times/week	32	43	10	27
	seldom	36	49	25	68
Topic knowledge	a lot	14	18	4	11
	some	48	61	27	71
	none	17	22	7	18

erably for different students, only one outlier had to be excluded from the analysis. T-test indicates that there is no significant difference between the two groups ($t = 0.65$, $df = 107$, $p = 0.520$). As table 2 shows, the average times spent by students in the Best-Seller group are marginally lower than the PACE group.

As figure 5 demonstrates, up to 63 percent of students in the Best-Seller group finished their searches within five minutes, while 46 percent of PACE group completed their searches within the same amount of time. After eight minutes, however, up to 80 percent of PACE group had stopped their searches, while only 71 percent of students in Best-Seller group had achieved their objectives. In general, students using Best-Seller had faster initial contact times than other students.

Retrieval Speed

Contact time may be interpreted in two different ways; users' frustration with the system and the inability to find useful information, or users' satisfaction with the system and the willingness to browse more than may be necessary for a specific search. Since either interpretation is valid, a better measure of the system performance and user satisfaction may be the time required to discover the very first call number or book. This time was also recorded by the research assistants, but not for all

Table 2
Contact Times and Retrieval Speeds

Group	n	Mean	s.d.	Median	Mode
Contact Time (minutes)					
PACE	71	6.2	4.6	5.2	5.3
Best-Seller	38	5.6	5.2	3.3	1.2
Retrieval Speed (minutes)					
PACE	67	2.7	1.7	2.3	1.4
Best-Seller	26	2.4	2.2	1.3	1.0

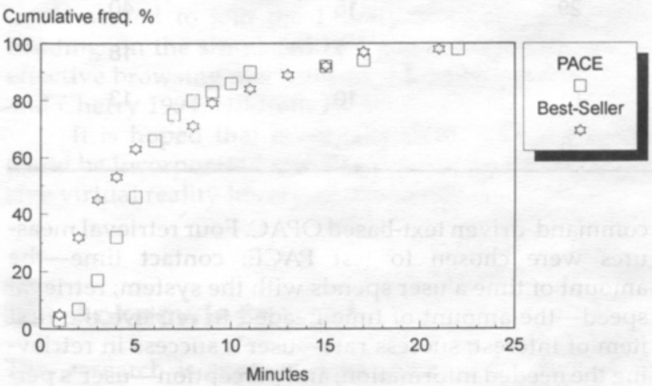


Figure 5
Contact time

the students. As table 2 shows, the average times for finding the first book are similar for both groups (t = 0.79, df = 91, p = 0.429). As figure 6 demonstrates, while up to 77 percent of students in the Best-Seller group were able to detect the first item of interest within three minutes, 64 percent of students in the PACE group were in this category. After five minutes, however, almost all the students in the two experimental groups had discovered the first title. It should be noted that students in the Best-Seller group had access to all the features of a powerful retrieval engine, such as Boolean and proximity operators, truncation, masking, and keyword searching in any field. PACE group was confined to using a mouse to browse relatively long lists and did not have access to a keyboard or all of the features of Best-Seller. Yet it is surprising that the majority of students in the latter group found their first book within three minutes, and their retrieval speed was not significantly different from the Best-Seller group.

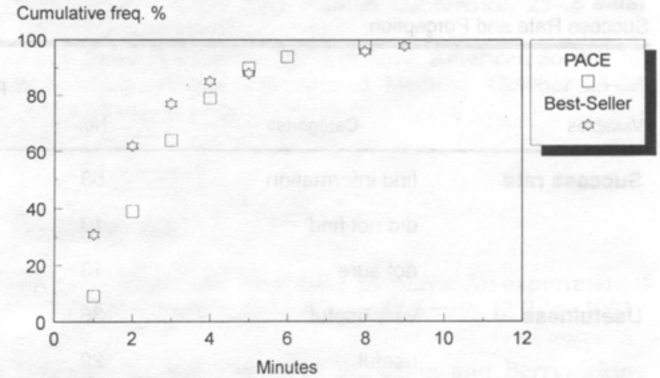


Figure 6
Retrieval speed

Success Rate

Students' success rate in finding materials for their information needs did not vary significantly between groups ($\chi^2 = 0.1039$, df = 2, p = 0.9494). About 67 percent of PACE group and 64 percent of Best-Seller group found materials pertinent to their needs (table 3).

Perceptions

Students' reactions were generally favorable toward both interfaces and they were perceived to be "very useful" or "useful," with no significant difference between the experimental groups ($\chi^2 = 0.0957$, df = 2, p = 0.9533).

Other factors measured to represent *perceptions* set PACE apart from Best-Seller. The two groups found the PACs easy to use. The PACE group, however, found this interface significantly easier to use ($\chi^2 = 12.1093$, df = 2, p = 0.0024). While only 42 percent of students in the Best-Seller group found their interface *very easy* to use, 69 percent decided that PACE had a *very easy* interface. At the other end of scale, while 18 percent of Best-Seller group found that interface not easy to use, only 3 percent of students in the PACE group were in this category. Other variables measured in this study do not have any significant effects on "ease of use" variable (p>0.05).

Perhaps the most important factor in choosing an interface for OPACs is users' preference. Students in the PACE group were asked about their preference for an interface. Since these students had already some experience with the Best-Seller interface, they could compare the two systems. An overwhelming majority (81 percent) chose PACE, with only 13 percent preferring Best-Seller and 6 percent undecided.

User Preference

An open-ended question was asked to solicit students' opinions about the experimental interface. Of those who

Table 3
Success Rate and Perception

Variables	Categories	PACE No.	PACE %	Best-Seller No.	Best-Seller %
Success rate	find information	53	67	25	64
	did not find	13	17	7	18
	not sure	13	17	7	18
Usefulness	very useful	36	48	18	46
	useful	29	39	15	39
	neutral/not useful	10	13	6	15
Ease of use	very easy	53	69	16	42
	easy	22	29	15	40
	neutral/difficult	2	3	7	18
Preference	prefer this system	63	81	10	13
	don't know	5	6		

volunteered their comments, all suggested that graphics and color were the best features of PACE. Students enjoyed the tidy and uncluttered screens of PACE, and over 95 percent liked the simulated bookshelves. Other positive comments involved the speed of the interface (84 percent), browsing the lists (75 percent), and the use of the mouse (73 percent). Students' preference for GUI-like features of PACE seems to conform with the results of another recent study on OPACs (Zorn and Marshall 1995). Some students indicated their preference for keyboarding rather than using only the mouse for searching PACE.

The look of the interface was not the only factor which determined user preference. Within the PACE group, preference was also closely associated with how easy ($\chi^2 = 8.0422$, df = 1, p = 0.0046) and how useful ($\chi^2 = 4.1480$, df = 1, p = 0.0417) students found that interface. Ten students did not prefer PACE to Best-Seller, yet nine found the interface *very easy* or *easy* to use. Preference is not dependent on other variables such as gender, topic knowledge domain, first language spoken, computer, OPAC, or library usage and year of study (p>0.05).

Discussion and Conclusion

PACE is designed to provide an additional option for novice users of online catalogs. In a real operational environment, PACE was tested against a traditional command-driven text-based OPAC. Four retrieval measures were chosen to test PACE: contact time—the amount of time a user spends with the system; retrieval speed—the amount of time needed to retrieve the first item of interest; success rate—user's success in retrieving the needed information; and perception—user's perception of the interface.

The results of this research show that when novice users are conducting subject searching in an online catalog, they can use a browsable graphical interface as effectively as a command-driven text-based system. No significant differences are observed between the two systems in terms of contact times and retrieval speeds. Although students in PACE groups were limited to using a mouse and did not have access to all the features of a second-generation OPAC —Boolean and proximity operators, keyword searching, etc.—they performed as well as the students in the Best-Seller group. The simple alphabetical scheme for browsing is an efficient approach for finding information. In fact, the mean contact time for the PACE group is slightly lower than the mean search time for another experimental interface with limited number of entries (Allen 1993). Success rate of students in retrieving the desired information was also similar between PACE and Best-Seller groups.

One of the most interesting results of this research is the overwhelming users' preference for a visual interface over a command-driven text-based system. Over 80 percent of students in the PACE group preferred it to Best-Seller. Perhaps, as Marchionini has suggested, students like browsing the simple lists provided in PACE because it requires a "smaller cognitive load" than using commands and other features of a traditional retrieval engine. Students found the simulated images of books and bookshelves in PACE intriguing and easy to browse through, as it is more intuitive to recognize information in a familiar metaphor than to recall it from memory. The cognitive burden imposed by PACE is far less than that of other systems. These results show that simulated images of books and bookshelves in conjunction with a simple browsable retrieval interface can provide a powerful search tool for novice users of OPACs.

Students' comments indicate the need for faster access mechanism for searching PACE. In the last phase of this project, which is currently in progress, a keyboarding feature is added to the retrieval system to increase the speed of access. A hyperlink system is also implemented to join the Library of Congress subject headings in the simulated title pages to facilitate more effective browsing mechanisms (Meadow, Marchionini, and Cherry 1994; Hildreth 1995).

It is hoped that eventually dynamic movements could be incorporated into PACE to create a nonimmersive virtual reality library environment.

Acknowledgment

This research is supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC) and Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR). The authors wish to thank Beverly Chandler, coordinator of the library at Vanier College, and Shirley Kearan, president of Best-Seller, and the staffs at both organizations for their generous support. We also like to thank our research assistants, Ann Harris and Margaret Cunningham.

Notes

1. It should be noted that the interface has undergone many revisions based on formal and informal feedbacks from the subjects in the first phase of the study and participants in different conferences where PACE was demonstrated. The conferences include:

2. *The Feasibility of a Browsable Interface for Online Catalogs*. American Society for Information Science 57th Annual Meeting, October 17–20, 1994. Alexandria, Va.

3. *The First PACE Towards a Virtual Reality Library*. Canadian Library Association 49th Annual Conference, June 14–18, 1994. Vancouver, B.C.

4. *Testing a Browsable Interface*. Canadian Association for Information Science 22nd Annual Conference, 25–27 May, 1994, Montreal, Quebec.

5. *Public Access Catalogue Extension*. American Society for Information Science 56th Annual Meeting, October 25–28, 1993, Columbus, Ohio.

References

Allen, B. "Improved Browsable Displays: An experimental Test." *Information Technology and Libraries* 12 (June 1993): 203–8.

Bates, M. J. "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." *Online Review* 13, no. 5 (1989): 407–24.

Beheshti, J. "Browsing Through Public Access Catalogs." *Information Technology and Libraries* 11, no. 3 (1992): 220–28.

Beheshti, J. Large, V., Bialek, M. "Testing a Browsable Interface, *Proceedings of the 22d Canadian Association for Information Science Annual Conference*. Montreal: Graduate School of Library and Information Studies, McGill University (1994).

Belkin, N. J., Marchetti, P. G., and Cool, C. "BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval." *Information Processing and Management* 29, no. 3 (1993): 325–44.

Borgman, C. L. "Why are Online Catalogs Hard to Use? Lessons Learned from Information Retrieval Studies." *Journal of the American Society for Information Science* 37 (1986): 387–400.

Borgman, C. L. "All Users of Information Retrieval Systems Are Not Created Equal: An Exploration Into Individual Differences." *Information Processing and Management* 25, no. 3 (1989): 237–51.

Borgman, C. L., Hirsh, S. G., and Walter, V. A. "Children's Searching Behavior on Browsing and Keyword Online Catalogs: The Science Library Catalog Project." *Journal of the American Society for Information Science* 46, no. 9 (1995): 663–84.

Buckland, M., Butler, M. H., and Norgard, B. A. "OASIS: A Front-End for Prototyping Catalog Enhancements." *Library Hi Tech* 10, no. 4 (1992): 7–22.

Busey, P., and Doerr, T. "Kid's Catalog: An Information Retrieval System for Children." *Journal of Youth Services in Libraries* 7 (1993): 77–84.

Chan, L. M. "The Library of Congress Classification System In An Online Environment." *Cataloguing and Classification Quarterly* 11, no. 1 (1990): 7–25.

Chang, Shan-Ju and Rice, R. E. "Browsing: A Multidimensional Framework." in *Annual Review of Information Science and Technology*, ed. Martha E. Williams. vol. 28. Medford, N. J.: Learned Information, Inc. (1993).

Chen, H., and Dhar, N. "Cognitive Process as a Basis for

Intelligent Retrieval Systems Design." *Information Processing and Management* 27, no. 5 (1991): 405–32.

Cochrane, P. A. *Improving LCSH for Use in Online Catalogs*. Littleton, Colo.: Libraries Unlimited, 1986.

Connaway, L. S., Budd, J. M. and Kochtanek, T. R. "An Investigation of the Use of An Online Catalog: User Characteristics and Transaction Log Analysis." *Library Resources & Technical Services* 39, no. 2 (1995): 142–52.

Cove, J. F., and Walsh, B. C. "Online Text Retrieval Via Browsing." *Information Processing and Management* 24, no. 1 (1988): 31–37.

Drabenstott, K. M., Demeyer, A. N., Gerckens, J., and Poe, D. T. "Analysis of a Bibliographic Database Enhanced With a Library Classification." *Library Resources and Technical Services* 34, no. 2(1990): 179–98.

Greenwood, D. "OPAC Research in the United Kingdom." in *OPACs and Beyond: Proceedings of a Joint Meeting of the British Library, DBMIST, and OCLC*. Dublin, Ohio: OCLC. 1989.

Hancock-Beaulieu, M., Robertson, S., and Neilson, C. *Evaluation of Online Catalogs: An Assessment of Methods*. British Library research paper 78 (1990). London: British Library Research and Development Department.

Hancock-Beaulieu, M., Fieldhouse, M., and Do, T. "An Evaluation of Interactive Query Expansion in an Online Library Catalog with a Graphical User Interface." *Journal of Documentation* 51, no. 3 (1995): 225–43.

Hildreth, C. R. *Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographic Retrieval Systems*. Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1989.

Hildreth, C. R. "The GUI OPAC: Approach With Caution." *The Public-Access Computer Systems Review* 6, no. 5 (1995): (http://info.lib.uh.edu/pr/v6/n5/hild6n5.html).

Lawrence, G. S. "System Features for Subject Access in the Online Catalog." *Library Resources & Technical Services* 29 (1985): 16–33.

Lee, N. S. "Multimedia Visualizer: An Animated Object-Based OPAC." *Information Technology and Libraries* 10, no. 4 (1991): 297–310.

Li, Tian-Zhu. "Generic Approach to CD-ROM Systems: A Formal Analysis of Search Capabilities and Ease of Use." In *Interfaces for Information Retrieval and Online Systems*, ed. M. Dillon. New York: Greenwood, 1991.

Marchionini, G. *Information Seeking in Electronic Environments*. Cambridge: Cambridge Univ. Pr., 1995.

Markey, K. *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs*. Dublin, Ohio: OCLC. 1984.

Meadow, C. T. *Text Information Retrieval Systems*. Toronto: Academic, 1992.

Meadow, C. T. Marchionini, G., and Cherry, J. M. "Speculations on the Measurement and Use of User Characteristics in Information Retrieval Experimentation." *The Canadian Journal of Information and Library Science* 19, no. 4 (1994): 1–22.

Michalak, T. J. "An Experiment in Enhancing Catalog Records at Carnegie Mellon University." *Library Hi Tech* 8, no. 3 (1990): 33–41.

Pejtersen, A. M. *The Book House: Modelling User's Needs and Search Strategies As a Basis for System Design*. Denmark: Riso National Laboratory, 1989.

Posey, E. D., and Erdmann, C. A. "An Online UNIX-Based Engineering Library Catalog: Purdue University Engineering Library." *Science and Technology Libraries* 6, no. 4 (1986): 31–43.

Poulter, A. "Towards A Virtual Reality Library." *Aslib Proceedings* 45, no. 1 (1993): 11–17.

Robertson, S. E., and Hancock-Beaulieu, M. "On the Evaluation of IR Systems." *Information Processing and Management* 28 (1992): 457–66.

Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 2d ed. (Ontario: Addison-Wesley, 1992.

Solomon, P. "Children's Information Retrieval Behavior: A Case Analysis of An OPAC." *Journal of the American Society for Information Science* 44 (1993): 245–64.

Tague-Sutcliffe, J., and Schultz, R. "Evaluation of the User Interface in an Information Retrieval System: A Model." *Information Processing and Management* 25, no. 4 (1989): 377–89.

Wahlde, B. V., and Schiller, N. "Creating the Virtual Library: Strategic Issues." In *The Virtual Library: Visions and Realities*, ed. Laverna M. Saunders, 15–46. Westport, Conn: Meckler, 1993).

Yee, M. M. "System Design and Cataloguing Meet the User: User Interfaces to Online Public Access Catalogs." *Journal of the American Society for Information Science* 42, no. 2 (1991): 78–98.

Zorn, M. J., and Marshall, L., "Graphical User Interfaces and Library Systems: End-User Reactions." *Special Libraries* 86, no. 1 (1995): 28–35.

At the Threshold of a Library Network

Farooq A. Khalid

The Arabian Gulf Libraries stand poised at the threshold of networking, which promises the sharing of resources through cooperation and provides the individual library with greater access to resources. Several libraries in the Gulf region have acquired sophisticated systems and there is a great potential to develop library networks. This paper highlights both the benefits and the problems associated with networking in libraries and discusses the circumstances that are forcing several information centers in the Gulf region to begin thinking seriously about library networking. The paper will briefly describe the development of an important networking tool, the Union List of Scientific and Technical Periodicals in the Gulf Region, and its current status and should serve as a springboard for further discussions related to library networks in the Gulf.

The control of knowledge is the crux of tomorrow's worldwide struggle for power in every human institution. **Alvin Toffler**

The past two decades have witnessed the formation of many library networks such as OCLC (Online Computer Library Center), RLIN (Research Libraries Information Network), WLN (Western Library Network), and many others. The developments in machine-readable standards for cataloging (MARC), computing, telecommunications, and the information industry have made it easier to develop cooperative programs such as library networks. The essence of this paper vis-à-vis library networks can be captured by two terms: *symbiosis* and *synergism*. Webster's defines *symbiosis* as a mutual cooperation between persons and groups in a society, especially when ecological interdependence is involved, and *synergism* as the cooperative action of discrete agencies (as drugs or muscles) such that the total effect is greater than the sum of the two or more effects taken independently.[1]

In order to benefit from library networks, network members must lose some independence as they must conform to the conditions laid out in the network contract. However, the advantages of networking far outweigh the disadvantages.

Definitions

In the literature one comes across several definitions of library networks. Susan Martin defines a network as a "group of individuals or organizations that are interconnected to form a system to accomplish some specified goal. This linkage must include a communications mechanism, and many networks exist for the express purpose of facilitating certain types of communication among members."[2]

Joseph Becker suggests that "when two or more libraries engage formally in a common pattern of information exchange, through communications, for some functionally interdependent purpose, we have a library network."[3]

Becker's definition is well suited to the context of this paper, as it is applicable to the emerging concept of library network in the Gulf region.

Components of a Library Network

Computer technology, communication technology, and information technology are necessary for the existence of library networks. The library network itself can be seen to have three dimensions: the network topology, the governing models, and the resource-sharing modules (see figure 1).

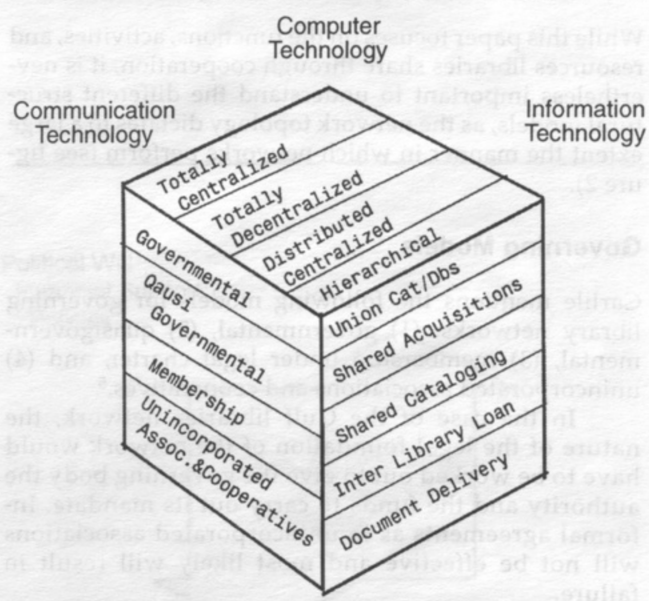


Figure 1
Three Dimensional Model of Library Networks

Farooq A. Khalid is Manager, Automated Information Systems Group, National Scientific and Technical Information Center, Kuwait Institute for Scientific Research.

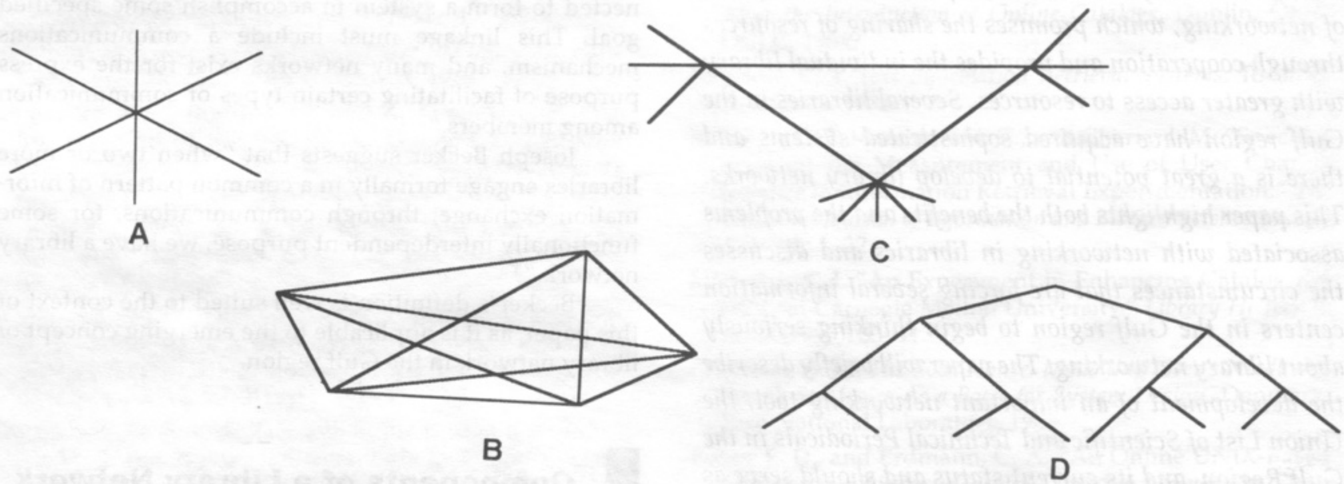


Figure 2
Topology of library networks as described by Williams and Flynn can be of the following types: (A) totally centralized (star) networks; (B) totally decentralized networks; (C) distributed centralized networks; (D) hierarchical networks.[4]

Network Topology

While this paper focuses on the functions, activities, and resources libraries share through cooperation, it is nevertheless important to understand the different structural models, as the network topology dictates to a large extent the manner in which networks perform (see figure 2).

Governing Models

Carlile mentions the following models for governing library networks: (1) governmental, (2) quasigovernmental, (3) membership under legal charter, and (4) unincorporated associations and cooperatives.[5]

In the case of the Gulf libraries network, the nature of the legal foundation of the network would have to be worked out to give the governing body the authority and the funds to carry out its mandate. Informal agreements as in unincorporated associations will not be effective and most likely will result in failure.

Resource Sharing

The essence of library networking lies in resource sharing through cooperative collection development, interlibrary loan, and document delivery policies that have been necessitated by prevailing conditions. Doughtery say that "in an effort to continue to provide access to the printed sources that still comprise the essential corpus of the world's scientific, humanistic, and social science literature, libraries have implicitly or explicitly endorsed resource sharing concepts."[6]

Why Should Gulf Libraries Network?

The limitation of financial resources coupled with the proliferation of information has placed library managers on the horns of a dilemma. This situation has forced many libraries throughout the world to look toward cooperative resource sharing through networking. The same is probably true for the libraries in the Gulf states, but because of poor standards and lack of cooperation, until now the formation of effective library networks in the Gulf has not been possible. The few networking success stories can be attributed to the persons involved rather than to the presence of a system.

In spite of the fact that today we in the Gulf can access a myriad of databases via online services and the Internet, indigenous materials of the Gulf states in the Arabic language remain elusive. To date there exists no bibliographic utility of Arabic materials within the Gulf countries; because there are no standards for cataloging Arabic materials, everyone duplicates the cataloging effort, producing records that are not shareable. From a networking viewpoint this is an exercise in futility.

The need for the development of bibliographic standards for Arabic materials is recognized by all, but a formal forum to tackle this problem has yet to be created.

Groen mentions that networks arise because the participating librarians understand that the individual library can no longer be self-sufficient. According to Groen, the most important and obvious ways in which to cooperate are through:

1. the development of an interlibrary loan or document delivery network;
2. the development of shared databases;
3. a cooperative sharing of expertise in answering users' reference questions;
4. the actual sharing of staff;
5. cooperative cataloging;
6. cooperative book and journal acquisitions programs; and
7. cooperation in the exchange of duplicate materials.[7]

In the case of the Gulf libraries, one could add the cooperation in the building of bibliographic standards and tools for Arabic materials.

The time is certainly ripe for networking in the Gulf, for the following reasons:

- We are all aware of the importance of standards and many libraries are implementing them in some form or another.
- Some major libraries in each of the Gulf countries have acquired sophisticated systems (VTLS being one of them) that will allow them to follow standards for bibliographic creations, holdings formats for serials, and the creation of authority files for subjects and names.
- The systems acquired also allow the creation of union catalogs.
- Gulf libraries now have a core group of experienced, qualified staff members.
- Dwindling financial resources are forcing us to look at networking alternatives.
- Whereas we can access materials published internationally, little accessibility to indigenous materials in Arabic exists.
- The progress in communications technology (e.g., the Internet) and the development of standard protocols (e.g., TCP/IP and Z39.50) have put networking within the reach of Gulf libraries.[8]
- Acceptance by the Library of Congress of the Arabic character set developed by the Research Libraries Group (RLG) (a project that was funded by the Kuwait Foundation for the Advancement of Science) will both serve as a standard for building Arabic databases and allow access to RLIN as a utility, a rich source of cataloged Arabic material.[9]

Most important, it must be realized that no one organization can forge ahead without the others. Networking is a joint venture and needs the cooperation of everyone involved.

Factors Responsible for the Success of Library Networks

Technological developments in informatics aside, the success of library networks can be viewed as dependent on the development from both the bottom-up and top-down processes. If the political will exists (at the top) but the basic technical processes do not (e.g., compliance with standards for cataloging, procedures for handling interlibrary loan and document delivery), the network will not be built. On the other hand, if the operational processes (at the bottom) are in place but the will among the decision makers is missing, then, too, the network will not see the light of the day (see figure 3).

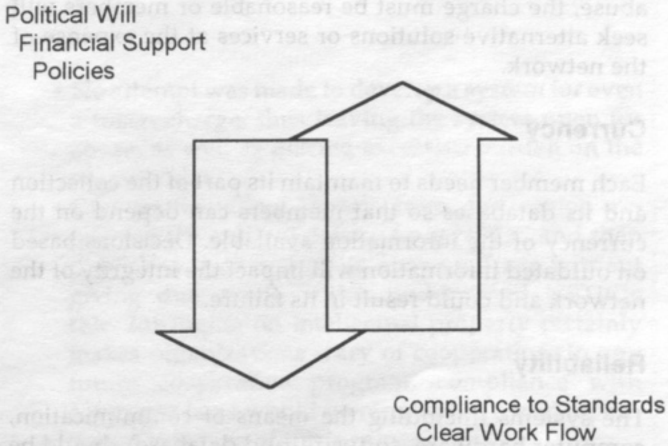


Figure 3

Other Factors: Interdependence, Timeliness, Cost-Effectiveness, Currency, Reliability

Other factors that influence the success or failure of library networks are:

Interdependence

Libraries join networks in order to increase their information base. Consequently, libraries with diverse collections but similar goals are better suited to network. Although specialization through shared collection development programs can ensure the enrichment of the information base, it also increases the level of interdependence of the members. The state of interdependence should not be jeopardized at any cost, because if it is, the network will fail.

Timeliness

Information exchange between libraries should take place in a timely manner. Delays can impair the effectiveness of the information if not render it totally useless. Through efficient workflow and improved methods and procedures, the viability of information exchanged between libraries can be guaranteed, which in turn will result in the success of the network.

Cost-Effectiveness

Whereas charging for a network service can deter its abuse, the charge must be reasonable or members will seek alternative solutions or services at the expense of the network.

Currency

Each member needs to maintain its part of the collection and its databases so that members can depend on the currency of the information available. Decisions based on outdated information will impact the integrity of the network and could result in its failure.

Reliability

The systems (including the means of communication, computer hardware, software, and database) should be well maintained and managed with reliable backups, and recovery and restart procedures. Frequent unavailability of even the most sophisticated system will result in the failure of the network.

Hindsight: The Union Catalog Project

As far back as 1976, The National Scientific and Technical Information Center (NSTIC) of the Kuwait Institute for Scientific Research (KISR) embarked on a project to publish a union list of scientific and technical periodicals in Kuwait representing thirteen libraries.

The idea behind this effort was to identify the periodical holdings of all the major libraries in Kuwait so that the process of resource sharing could begin. The success of the union list in Kuwait encouraged NSTIC to include periodical holdings from major libraries in the Gulf. As a result, in 1977 a regional union list was published that included eleven libraries and twelve branches from Iraq and three libraries with twenty branches from Saudi Arabia and encompassed twenty-eight subject categories. In 1980 NSTIC also undertook the publishing of a union list of Arabic periodicals in Kuwait.

In 1980 the regional union list was expanded to include sixteen libraries from Kuwait and a library from the United Arab Emirates. In 1983 the Regional Union List was published in a COM (computer output on microfilm) format.

It was realized that an online database of the regional union list would be effective and more meaningful if the data were in the MARC format. This sentiment is reflected in the concluding remarks of Howard's article titled *The Future of the Union Catalog* in which he states "the main thrust of standardization is to permit sharing and cooperation among libraries. The purpose of cooperation is to facilitate the connection of person and information. It is hoped that the new NUC will be a useful medium for that connection."[10] To this end, NSTIC started to convert the entire union catalog of serials database conforming to MARC.

Having lost everything in the invasion of Kuwait, NSTIC started from scratch, and to date it has developed a union catalog of journal holdings of Kuwait University, Kuwait Fund, and the Arab Planning Institute in MARC format.

Observations Related to the Union List Project

Listed below are some observations related to the development of the regional union list of scientific and technical periodicals carried out by NSTIC.

- In the years that the union list was published, the entire responsibility was borne by NSTIC. The greatest degree of cooperation offered by the participating libraries lay in making their catalogs

Milestones of the Union Catalogs Project in NSTIC

1976	Union List of scientific and technical periodicals in Kuwait libraries and approximately 2,180 journal titles published jointly by NSTIC and Kuwait University (KU).
1977	Regional Union List (Including thirteen libraries from Kuwait, eleven from Iraq, and three from Saudi Arabia and twenty-eight subject categories in the fields of science and technology) published jointly between NSTIC and KU. KU contributed half of the printing costs.
1977	Updated Kuwait Union List of scientific and technical periodicals.
1978	Updated Kuwait Union List of scientific and technical periodicals.
1980	Kuwait Union List of Scientific and Technical Arabic Periodicals representing ten libraries and approximately 600 titles. Published by NSTIC.
1980	Regional Union List (English) expanded to sixteen libraries from Kuwait and one library from the UAE. The list had approximately 7,888 journal titles.
1981	National Book Catalog (including eight libraries and approximately 50,096 book titles).
1981	Updated Arabic Union List to include holdings of twelve libraries from Kuwait. The list had approximately 700 titles.
1983	Regional Union List (English) updated and converted to COM (computer output on microfiche) format.
1983–1992	Converted all Kuwait journal titles to MARC format.
1992–present	Union Database of Journal titles of NSTIC, Kuwait University, Kuwait Fund, and Arab Planning Institute in MARC format for approximately 5,000 titles.

available for NSTIC data-capturing staff. Kuwait University offered to pay half the printing costs one time.

- Even after many years of use, no formal cooperative interlibrary loan policy emerged, which should have been a natural consequence of the product.
- It was initially expected (perhaps prematurely) that eventually every participating library would start to code its own updates, leaving NSTIC to collate, sort, and update the database; however, no formal decentralized updating methodology has been agreed upon.
- Some document delivery (e.g., photocopying of articles) was exchanged between libraries, but again no concerted effort was made to follow up and formalize the procedures, thereby undermining the reliability of the document delivery procedures.
- No attempt was made to develop a system for even a token charge, thus leaving the system open for abuse, as well as placing excessive burden on the net provider.
- One participating library from the region requested the entire database from NSTIC and then published it as one of its own products without giving due credit, which undermined NSTIC's role. Infringing on intellectual property certainly makes organizations wary of cooperation. In any future cooperative program, compliance with copyright laws would be required to avoid repetition of this type of infringement.
- As the developer and producer of the union list database, NSTIC for its part could have been better organized, making the project a greater success (see sidebar above).

Concluding Remarks

Networks arise when individual libraries realize that they can no longer be self-sufficient. The underlying assumption of a library network is that all the members will share their resources and use the network. Cooperation is the key word. Without it, the network will fail.

The issue of access versus acquisitions pops up when networks are discussed. Ideally every library would like to acquire all its information needs, but because of the explosion of information and limited resources, library networking appears to be the only practical solution available.

The libraries in the Gulf should grab this opportunity and not delay in their decision to network. The climate is right. Communication technology, standards, and protocols are in place, and computer systems and bibliographic databases exist. The next logical step for the Gulf libraries is to create a formal library network. The effectiveness with which networks can be established depends on everyone involved—technocrats, administrators, and individual librarians. The fruits of synergy are waiting to be realized.

References

1. *Webster's Third New International Dictionary*, Merriam-Webster Inc. s.v. "networking."

2. S. K. Martin, *Library Networks, 1986–87: Libraries in Partnership* (White Plains, N.Y.: Knowledge Industry, 1986).

3. J. Becker, "Network Functions: Reactions," in *The Structure and Governance of Library Networks: Proceedings of the 1978 Conference in Pittsburgh*, ed. A. Kent and T. J. Galvin, (New York: Marcel and Deckker, 1979), 85–91.

4. J. G. Williams and R. Flynn, *The Structure and Governance of Library Networks: Proceedings of the 1978 Conference in Pittsburgh*, ed. A. Kent and T. J. Galvin (New York: Marcel Dekker, 1979), 49–83.

5. H. Carlile, "The Diversity among Legal Structures of Library Networks," in *Networks for Networkers*, ed. B. E. Markuson and B. Wolls (New York: Neal-Schuman, 1980), 187–210.

6. Richard M. Dougherty, "Resource Sharing among Research Libraries: How It Ought To Work," *Collection Management* 9, nos. 2/3 (Summer/Fall 1987): 79–111.

7. F. Groen, "A Comparative Review of Medical Library Networking in Canada and the United States," *Health Information and Libraries* 1, no. 3 (1991): 111–18.

8. L. Eppelheimer, "Information Retrieval: A Unified Format," IOLS '94, *Integrated Online Library Systems: Proceedings, 1994, New York, May 11–12, 1994* (Medford, N.J.: Learned Information, Inc., 1994).

9. Joan M. Aliprand, "Arabic Script on RLIN," *Library Hi Tech* 40 10, no. 4 (1992): 59–80.

10. J. H. Howard, *The Future of the Union Catalog: The Library of Congress Union Catalog Experience*, International Symposium on the Future of the Union Catalog, University of Toronto, 1981 (New York: Haworth, 1982).

Digital Imaging on a Shoestring: A Primer for Librarians

Susan Alden

In April 1994 Cline Library at Northern Arizona University (NAU) received funding for a pilot project to establish a database of digitized original photographs, manuscripts, and maps from the Special Collections and Archives Department. Materials were selected for digitizing to support curriculum and instructional needs of four NAU faculty whose classes were broadcast to distance sites statewide during Fall 1995. PCs for accessing the image database were installed at four distance sites, and library staff at these sites were trained in the use of Netscape for database access and navigation. The pilot project has provided an opportunity for close collaboration of library staff with other NAU staff working in the areas of distance education, instructional technology, and network communications.

Why Digitize?

Thousands of institutions worldwide are committing substantial resources to digitizing collections. Is it worth it? What are the benefits? A primary benefit is to provide remote access to library and archival materials. Online catalogs identify and provide references to material. Image databases provide access to digital representations of the materials themselves, a distinct advantage to the user.[1] Because in the digital world, preservation and access become synonymous, a second important benefit of digital technology is that it provides a preservation format that supplements traditional preservation practices. The Commission on Preservation and Access defines preservation as "the provision of access to knowledge recorded on a multitude of media as far into the future as possible."[2]

Northern Arizona University has been given a specific mission by the Arizona Board of Regents: to serve the communities of rural Arizona by making its resources available to them using appropriate technology. The Special Collections and Archives Department at Cline Library houses a photograph collection of 700,000 images, a manuscript collection of 3 million documents and a map collection of 1,300 documents. These collections depict the history and development of the Colorado Plateau from prehistory to present and contain materials related to forestry, ranching, mining, geology, archaeology, Native American traditions, urban settlement and growth, tourism, and recreation. All materials at Special Collections and Archives are noncirculating. For place-bound students at over fifty NAU distance sites around Arizona, these materials remain inaccessible due to geographic distance.

Critical Decisions To Be Made

Many important decisions need to be made before establishing an image database. Those decisions will affect how the database is accessed and navigated by local and remote users, who those users will be, and the future viability of the database as technologies change.

Academic Support

Will the database benefit a broad range of academic departments or only a few? Will faculty have an interest in using the image database as an instructional tool? Will they need assistance incorporating it into the curriculum? New technologies are embraced more wholeheartedly by some faculty than others, depending on teaching methods, the nature of the discipline being taught, and personal preferences.

Funding

How much is needed and from what source? Creating an image database can be very costly, depending on the technologies you choose and the level of technical support your institution provides. Our pilot project was funded by the university. We're investigating grant sources for future funding.

Technologies

What kinds are most appropriate for your institution? Will all components be compatible? Who will train your staff to use them? Even inhouse technical support staff may not be familiar with imaging technologies.

Data Standards

Which file formats are most appropriate for your data? Should data files be compressed? Which compression method is best? It is vital

Susan Alden is Records Retention and Information Technology Manager, State of Washington, Office of the Secretary of State, Archives and Records Management Division, 1120 Washington St. SE Olympia, WA 98504-0238. At the time of the project, the author was Program Coordinator at Cline Library, Special Collections and Archives Department, Northern Arizona University. The author wishes to thank T. G. McFadden for his editorial assistance in preparing this article and Tamara Lopez for her continuing inspiration and creativity throughout the imaging project.

to select data standards that not only are widely used now but are likely to be supported by future technologies.

Indexing and Retrieval

Database users depend completely on effective indexing and a good search engine to navigate your image database. Will you fully catalog materials or provide more brief descriptive records? Will you describe materials at the collection or item level?

Security

Ease of access to your database is both its strength and its liability. Anything on the Internet can be downloaded by people with the technology and skills.

Copyright

Your institution could be legally liable for making resources Internet-accessible if you do not hold the copyright or if the resources are not in the public domain as defined by copyright law.

The Pilot Imaging Project at Cline Library

Two databases of digitized materials have been created during the pilot project: an archival database with controlled, local access and a browsing database on the World Wide Web. The archival database resides on a local hard drive and consists of catalog records in MARC format with associated high-resolution image files. The catalog records were created using software developed by a volunteer programmer working on the project. Image files were created by scanning materials at 600 pixels per inch (ppi) using an Arcus II color flatbed scanner. Image files in the archival database are stored in TIFF (Tagged Image File Format) format using LZW, a lossless compression method that does not degrade the image when the file is decompressed. We developed our imaging and database maintenance procedures by combining expertise from local imaging and systems people with a great deal of on-site experimentation.

Both catalog records and image files are modified for display in the browsing database on the Web. The software used to created the archival database strips the numeric MARC tags from the catalog record, displaying simple field names to describe the data. The software converts the catalog records to hypertext markup language (HTML) format and embeds links to associated image files. Image files are converted to low resolution (72 ppi) and GIF (Graphic Image File) format for Web display using Adobe Photoshop software. Adobe Photoshop is also used for image enhancement.

File format conversion can degrade image quality. Photographs with good tonal contrast generally produce good-quality scans, but images with flat tonal range and those that are underexposed or overexposed often require substantial enhancement for Web display. Digital images of maps and manuscripts generally require some reworking to highlight text and minimize the effect of textured backgrounds.

We kept the imaging project technology as simple as possible. Rather than purchasing state-of-the-art equipment and software, we selected well-established technologies with proven performance, supplementing the basic configuration with the extra power and speed required to process an imaging workflow efficiently. The main expense beyond the basic equipment was the additional storage capacity required for the archival database. We purchased a 10 gigabyte hard drive initially and found we needed an additional 10 gigabytes within the first year.

The image database uses Glimpse search software, developed at the University of Arizona and modified for project use by Computing Technology Services staff at NAU. Glimpse accommodates Boolean and wildcard keyword searching and has an optional form-based interface that allows the user to make temporary changes to several defaults: case sensitivity, the number of misspellings allowed, the maximum number of files returned, and the maximum number of matches per file. The search query result consists of document titles with hot links to the corresponding catalog records.

At the beginning of the project we initiated discussions with faculty teaching in disciplines that are a focus of the Special Collections and Archives Department, including the Colorado Plateau and the history of the Southwest. We also met with faculty who were active in the use of instructional technology in the classroom. For this pilot project we worked closely with four faculty members whose classes were broadcast statewide via interactive television on selection of materials for digitizing to support their curriculum and instructional needs. All participating faculty members were trained in access and navigation of the database using the Netscape browser.

The database currently consists of about 700 catalog records with associated image files. Digitized portions of collections with significant research value are being added to the image database, which will reach 1,000 catalog records within the next six months. Cline Library's online catalog will contain summary catalog records for digitized collections, with a hot link to the image database on the Web.

The image database resides on the World Wide Web at http://www.nau.edu/~cline/speccoll/imagedb.html.

Some Things We've Learned

Pilot projects are a wonderful opportunity to learn by doing. We've faced many decisions in technical areas that we didn't know even existed before beginning the project. Here are some ideas to make your learning curve less steep.

Network with Others

Speak with others who have digitized collections to find out what their experiences have been. You can join an Internet discussion group or form a regional focus group. The Arizona State Imaging Network is composed of academic and cultural institutions in the process of digitizing collections or considering digitization projects. We have a wonderful exchange of information and learn a great deal from each other. You can also hire a consultant to help you define the specifics of your database project. Keep the following guidelines in mind when you're planning your project:

Start Small

Keep it small and simple until you've assessed how it's working. It's much easier to make revisions with a limited number of records in your database.

Establish a Timetable, but Expect the Unexpected!

Pilot projects that use emerging technologies are by their nature unpredictable. We've worked through each challenge using teamwork, networking, and systematic problem solving.

Ensure That Your Project Has Adequate Systems Support

In most institutions systems support people are overextended, trying to keep everyone's PC working. We relied on in-house systems staff for installation, configuration, and maintenance of the imaging project equipment, and they have done an excellent job of providing support. It is best to have a systems support person working directly with your project. However, in order to address the specific needs and challenges that arise from the complexities of imaging equipment and software, the scanning process, and the maintenance and storage of large image files, be sure your project also has the administrative support to make it a high priority.

Develop Methods to Assess What's Working

This involves examining the procedures used to *create* and to *access* the database. Survey remote users and get their comments on how database access and navigation is working for them.

Be Willing to Change How You're Doing Things

It may have seemed like a great idea when the project began, but is it working now? Continually reassess your methods. We've changed some procedures four or five times. The way we do things now reflects what we've learned.

Think in the Future Sense

How will this database be accessed and navigated in the future? Your online data will need to be migrated to future systems, with different search software and storage devices. What provisions have you made for future data migration?

Trust in the Process of Teamwork

We began the imaging project with a $15,000 budget for in-house equipment, including the imaging workstation, scanner, and storage devices for the local and remote systems. None of us had previous experience with any aspect of imaging. We knew what we wanted to accomplish, but not how to do it. We worked through the uncertainty by trusting the process of teamwork. Our team has included not only staff at Cline Library but people in other departments and at other institutions who provided the skills and information we needed to make many critical decisions. We found that people were quite willing to share their time and expertise and became as enthusiastic about the project as we were. These resources are available to your institution as well. The Cline Library staff encourage colleagues to consider digitizing selected collections for Internet access—you really *don't* need grant money to do it!

References

1. *The Commission on Preservation and Access, Annual Report, 1991–1992* (Washington, D.C.: The Commission, 1992).

2. Clifford A. Lynch and Edwin B. Brownrigg, "Conservation, Preservation, and Digitization," in *Association of College and Research Libraries, National Conference, 4th: 1986: Baltimore, Energies for Transition*. Association of College and Research Libraries: 225–28.

Selected Bibliography

Besser, Howard. "Getting the Picture on Image Databases: The Basics." *Database* 18 (Apr.–May 1995): 12–19.

Conway, Paul. "Digitizing Preservation." *Library Journal* 119, no. 2 (Feb. 1994): 42–45.

Gonthier, Dominique. "The Digital World and Intellectual Property. Opportunities and Challenges: The Technological Answer." *Information Services & Use* 14, no. 4 (1994): 237–42.

Hamber, Anthony. "The Challenge of Digital Imaging Technologies: A Practical View of the Future." *Information Services & Use* 14, no. 3 (1994): 243–51.

Lynch, Clifford A. "The Integrity of Digital Information: Mechanics and Definitional Issues." *Journal of the American Society for Information Science* 45 (Dec. 1994): 737–44.

Quinn, Richard A. "Networked Document Imaging Software: Guidelines to Providing Widespread Access to Information." *Electronic Library* 12 (Dec. 1994): 370–71.

Local Holdings Searching in ISI® CD-ROM Databases

Charles F. Priore, Jr., and Bryn Geffert

Rationale

St. Olaf College is proud of its chemistry department, graduating more chemistry majors during the last decade than any other four-year liberal arts college in the United States. Given the number of chemistry majors and the amount of chemical research conducted at St. Olaf, it is not surprising that Chemistry Citation Index® (CCI)—published by the Institute for Scientific Information (ISI) and one of the primary indices to chemistry journals—receives heavy use. Students, ranging from upperclassmen assisting their professors with original research to underclassmen just beginning to explore the chemical literature, rely heavily upon this index.

The St. Olaf Science Library subscribes to only a small portion of the many journals indexed by CCI. Students and faculty who need exhaustive searches of the chemical literature find CCI invaluable and can use Minitex, the library network funded by the state of Minnesota, to obtain articles not available inhouse. But for chemistry students just beginning to familiarize themselves with the literature, CCI can be a frustrating product when only a small number of the citations retrieved are available on campus.

To address this predicament, the St. Olaf Science Library devised a method of limiting CCI searches to in-house titles. Creating a local serials holdings list of Silver Platter CD-ROM databases has already been demonstrated,[1] but the customization of ISI CD-ROM products proved to be quite a challenge. Nevertheless, with a little ingenuity we were able to incorporate a local serials list into CCI, devising a method that is also applicable to other ISI products.

Students find this local holdings feature invaluable, allowing them to use one of the primary indexes to the chemistry literature while ensuring their time is not wasted wading through citations to journals owned only by large or specialized research institutions.

Methodology

The following instructions are for version 3.05 of the ISI search software.

We began with St. Olaf's locally created serials list and with a printout of titles representing monographic serials such as the *Annual Review of Biochemistry*, etc. After powering up the computer and inserting the CCI CD into the CD-ROM drive, you are defaulted to the "basic index" search option. Perform an Alt-F (fields) command. Using the arrow keys, key down to the "abbr Journal" option (in some versions of the software you may see "abbr sOurce" rather than "abbr Journal"). Perform an Alt-D (Dictionary) command. You will now see CCI journal titles in their abbreviated form. Using the space bar, mark the titles to which your institution subscribes. Using the down arrow key, continue marking titles, making sure that you do not exceed the space buffer

Charles F. Priore, Jr., is Associate Librarian (Natural Sciences), St. Olaf College and Carleton College, Northfield, Minnesota. **Bryn Geffert** is Assistant Professor (Reference and Bibliographic Instruction Librarian), St. Olaf College.

in the "abbr Journal" field (you will receive the message "search query exceeded maximum length" should you overload the buffer). After you have selected several titles, hit the Enter key. Press Enter again and the titles will be searched, causing a "Total" to appear in the "Records Search Terms" window. Press Enter yet again and "Set 1" is created. Perform another Alt-D; scan down the list using the space bar to mark your selections. Hit the Enter key three times, and "Set 2" is created, and so on.

When all journal titles have been selected it is time to OR together all of the search sets. This time choose the Alt-F (fields) option. Now choose "S" for set combinations. Begin entering the set numbers (e.g., 1 or 2 or 3 or 4), once again making certain the strategy does not exceed the space buffer. Once all sets have been OR'd together the final set number will contain all serials to which the home institution subscribes.

Now perform an Alt-S (Save Strategy) command. You will be prompted to "save search strategy to: C:\(name file)." After naming the file (we use JNLSCHEM), hit the Enter key and the search strategy will be saved to your hard disk.

A student can now use the saved file to limit searches to local holdings. Press Alt-R (Run Strategy option). The computer will begin searching the highlighted titles and creating sets. The last set number of the search session is the number the student will use to locate local holdings. For example, the student performs an Alt-R and the last set number is 145. After performing a subject search the student will use the AND operator to combine the subject search set with the last set number in the "Strategy" session. Using the "Set Combination" option in the Alt-F menu, the student would type the command: "145 and 149," where 145 represents the total from the serials search and 149 represents the last set containing citations from the subject search. The resulting AND search produces citations that can be found at the home institution.

Requirements

The ISI manual calls for at least an IBM XT (manufactured by DTK) or AT to run the search software. We ran our searches on two different 486 PCs (DTK), both using a NEC CD-ROM drive and version 2.3 or version 2.31 driver software. The ISI manual calls for at least 640K of RAM, and we ran a successful search with just over the bare minimum. The manual also calls for a minimum of 3 free megabytes hard disk storage. We were able, however, to run the search successfully with less than 2.8 megabytes. Free disk space should not be a significant issue; our title file took up relatively little room, consisting of 11,000 bytes.

Performance

The run time for the journal file on a 486 computer with 2,707,000 bytes free disk space and 543,120 bytes free RAM was one minute and fourteen seconds. One should expect the run times on lower end machines to be significantly longer.

Conclusion

The local holdings file has proven very popular with our students and faculty. Although impossible to prove quantifiably, interlibrary loan use has lessened considerably as a result of students' ability to limit searching to in-house journals. Faculty encourage students to use the local holdings feature, and the frustration evident before this capability was implemented has greatly diminished.

The process, however, is not without its difficulties. Large universities and research institutions with extensive journal holdings would exceed the amount of RAM required to conduct a local holdings search (although limiting searches to a small set of titles is less of a concern for such institutions). The holdings file must be regularly maintained, canceled journals must be purged, and when initially creating the holdings list one must be careful not to enter titles canceled before the date of database coverage. It is impossible to specify dates of owned journals, a shortcoming that may cause some confusion.

Note

1. Charles F. Priore, Jr., and Richard E. Miller, "Local Holdings Searching in CD-ROM Databases," *Information Technology and Libraries* 11, no. 3 (Sept. 1992): 307–9.

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION

Information Technology and Libraries, Publication No. 280-800, is published quarterly by the Library and Technology Association, American Library Association, 50 E. Huron St., Chicago (Cook), Illinois 60611-2795. The editor is James J. Kopp, Library Director, Wilson W. Clark Memorial Library, University of Portland, P.O. Box 83017, 5000 N. Williamette Blvd., Portland, OR 97283-0017; the managing editor is Ann Leslie Jones, Ryerson and Burnham Libraries, The Art Institute of Chicago, 111 S. Michigan Ave., Chicago, 60603. Annual subscription price, $25. Printed in U.S.A. with periodical-class postage paid at Chicago, Illinois and other locations. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 only), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

EXTENT AND NATURE OF CIRCULATION

(Average figures denote the average number of copies printed each issue during the preceding twelve months; actual figures denote actual number of copies of single issue published nearest to filing date: June 1996 issue.) Total number of copies printed: average, 7,598; actual, 7,380. Sales through dealers, carriers, street vendors and counter sales: none. Mail subscription: average, 6,896; actual, 6,728. Free distribution: average, 90; actual, 88. Total distribution: average, 6,986; actual, 6,816. Office use, leftover, unaccounted, spoiled after printing: average, 612; actual, 564. Total: average, 7,598; actual, 7,380. Percentage paid: average, 98.71; actual, 98.71.

Statement of Ownership, Management, and Circulation (PS Form 3526, Oct. 1994) for 1995/96 filed with the United States Post Office Postmaster in Chicago, October 1, 1996.

News and Announcements

LITA Awards

Several awards were announced and presented at the LITA President's Program during the ALA Annual Conference in New York this past July:

LITA Library Achievement Award

Paul Sybrowsky and D. Keith Wilson, cofounders of Dynix, Inc., won the 1996 LITA Achievement Award in Library and Information Technology. The award, a citation and $1,000, recognizes achievement in library and information technology in the areas of leadership, development of a technical application, accomplishments in research or education, or a contribution to literature in the field.

LITA Library HiTech Award

The Continuing Education Services Program at the School of Library and Information Studies at the University of Wisconsin-Madison won the 1996 LITA Library HiTech Award. The award, a citation and $1,000, recognizes an individual or institution for a single, seminal work, or body of work, taking place within the five years preceding the award that shows achievement in communicating to educate library practitioners in library and information technology.

LITA/LSSI Scholarship

Maurice Okereke won the 1996 LITA/LSSI Minority Scholarship in Library and Information Technology. Mr. Okereke is the long-term school librarian at Girls' High School in the Philadelphia school district. The scholarship consists of a $2,500 stipend provided by LSSI.

LITA/OCLC Scholarship

Yin-Fen Pao won the 1996 LITA/OCLC Minority Scholarship Award. Pao is enrolled in the School of Library and Information Science at Wayne State University. The scholarship consists of a $2,500 stipend provided by OCLC.

LITA/GEAC Scholarship

Angela Hodge won the 1996 LITA/GEAC Scholarship in Library and Information Technology. Hodge is enrolled in the School of Library Science at Emporia State University. The scholarship consists of a $2,500 stipend provided by Geac Inc. For further information on the LITA awards and scholarships, contact the LITA office, ALA, 50 E. Huron St., Chicago, IL 60611; 1-800-545-2433, ext. 4269; e-mail vedmonds@ala.org.

Future of Higher Education Networking

In August 1996 the Coalition for Networked Information (CNI) hosted a workshop with Advanced Network & Services, EDUCOM NTTF, FARNET, and NYSERNet (with support from the RAC and the NSF) on the future of higher education networking. Over 100 participants from higher education, the research community, government, and commercial sectors came together to organize thinking and effort regarding how best to meet the requirements for the next generation of networking in the higher education community.

The proceedings from the workshop can be found at www.farnet.org/cheenne. Several presentations were made at the workshop, and those slides are available to view online or to download as PowerPoint files.

Digital Object Identifier System

The Association Of American Publishers (AAP) announced in September the selection of the team that will develop a Digital Object Identifier (DOI) system for use by the publishing industry. The team, composed of R.R. Bowker and the Corporation for National Research Initiatives (CNRI), was chosen following a competitive bidding process initiated last spring as part of AAP's all-out effort to promote development of systems for managing copyright in the digital environment. The DOI project is the outgrowth of a year-long aap initiative to identify the needs of the publishing industry to facilitate safe and successful commercial ventures on the internet and in other networked environments. aap's initial research revealed a fundamental need for an electronic "license plate" for a "digital vehicle" traveling the information superhighway. The digital object identifier system will serve that purpose. For more information, contact Maureen Adamson, vice president of business development at Bowker, at (908) 665-2856 or madamson@reedref.com, Or Constance Mclindon, director of system development at CNRI, at (703) 620-8990 or Mclindon@cnri.reston.va.us.

Index to Advertisers

Information Technology and Libraries
Index to Volume 15 (1996)

Compiled by Kieth C. Wright

Set your information free!

Introducing Minolta EPIC 3000

Scan any bound document face up and damage-free in seconds.

IMAGING MAGAZINE PRODUCT OF THE YEAR 95

Scan, copy and send information from any bound document interoffice, interlibrary or internationally.

Bound up by bound documents? Digitize and distribute information worldwide with one of the EPIC 3000 systems. Scan fragile books, documents and even 3-D objects face up and damage-free. Then archive or distribute the information however you please. EPIC 3000 for Windows™ lets you print or fax. E-mail information. Store it on any form of media. Or put it on the Internet.

The patron-friendly EPIC 3000 with touch screen comes ready for paper or electronic output. Don't get caught in a bound document bind.

For more information call *1-800-9-MINOLTA*

Simple touch screen monitor lets patrons print or store to disk.

Copiers
Cameras
Faxes
Digital Systems
Document Imaging
Camcorders
Binoculars
Color Sensors

Only from the mind of Minolta

MINOLTA