Seeking the Subject*

JENNIFER TOBIAS

ABSTRACT
As an exercise in pondering cataloging in the networked environment ("networked environment" meaning electronic information sources interconnected through the Internet), this article compares traditional cataloging and Web-based description of two topics—the concept of "green cards" and a recent nonfiction work. This involves, first, outlining intellectual access issues as they apply to reference services today ("intellectual access" meaning the formal or informal description of a work for purposes of its discovery by others). Following this is an outline of key cataloging issues, per Sanford Berman, and corresponding issues in Web-based intellectual access. Ways that catalogers and public service librarians can address these issues conclude the article.

THE REFERENCE SCENE
Several key issues in intellectual access apply to reference services today. Perhaps the most crucial is an increasing demand for what this author calls "naïve" access—i.e., access to specialized subject knowledge by nonspecialists in that subject. Two major trends contribute to this demand. The first is the sheer volume of scholarly, professional, and popular publication. The second is a general intellectual trend toward interdisciplinarity (for one perspective on the implications of interdisciplinarity, see Messer-Davidow, Shumway, & Sylvan, 1993).

* Excerpted from "Issues of Intellectual Access in Our Electronic Age," with Elliott Shore (Director of Libraries, Bryn Mawr College) and Sanford Berman (Chief Cataloger, Hennepin County Library, Minnesota) a presentation of the Rutgers University SCILS Professional Development Program, April 1996.

Jennifer Tobias, Reference Department, Library, Museum of Modern Art, 11 West 53rd Street, New York, NY 10019
© 1998 The Board of Trustees, University of Illinois
As a result, librarians must be fluent in many subject vocabularies. Consider, for example, a historian seeking technical information about medical effects of lead use in ancient material culture. What vocabularies apply (historical, medical, chemical, sociological, or material)? The librarian must be able to communicate across these disciplinary vocabularies. This is a crucial skill for reference librarians today.

A third key issue is that increasing sophistication of hypermedia creates higher expectations by information seekers. In concrete terms, if home-Internet and twenty-four hour news channels (not to mention cars, phones, and the occasional coffee maker) appear to respond so readily to our everyday information needs, why is it so hard to pursue a question at the library?

This sophistication is also redefining what constitutes a scholarly work. Contemporary scholarship in the humanities and social sciences, for example, more often looks to the medium of messages—i.e., the way ideas are conveyed as much as the ideas themselves (see this line of thinking applied to hypertext navigation in Aarseth, 1997, chap. 8). The rise of media studies and the methodology of deconstruction are but two examples of this.

Further, the products of such scholarship are increasingly likely to be expressed in multiple media. In the humanities, a good example of this is the Perseus Project (http://www.perseus.tufts.edu), a hypermedia work thoughtfully integrating history, geography, literature, cultural studies, material culture, and mythology. In the "hard" sciences, as computing becomes increasingly integrated into methodology, the results of research increasingly integrate computing. Think of the Human Genome Project (http://www.nhgri.nih.gov/HGP), the international gene-mapping collaboration.

Finally, fast-paced changes in information technology are having obvious effects upon reference services. The task of integrating reference resources in diverse formats is one. The breakdown of distinctions between reference services and computing/information services is another. Ways to address these issues are discussed in the concluding section.

THE CATALOGING SCENE

To outline some key issues in subject cataloging, this discussion will now turn to the indefatigable Sanford Berman (1993) and summarize his longstanding critique of traditional cataloging—i.e., deficiencies in traditional (AACRII) cataloging conceals works. These deficiencies are illustrated by a search in a traditional catalog (the Library of Congress catalog is used here) for information about "green cards" (representing "resident alien" immigration status in the United States). These deficiencies include:
• Anachronistic subject headings (A previous heading in LCSH was *Alien registration receipt cards (United States)*)
• Table-of-contents absent
• Lack of added titles
• Lack of notes
• Poor cross-referencing (for comparison, see the Hennepin County Library's treatment of the topic, in particular the scope notes visible in the catalog).

Compare this to indexing of this topic with the Yahoo! Internet search engine (http://www.yahoo.com) (for purposes of argument, the dominant advertisements for immigration lawyers have been ignored). The result includes the site illustrated in Figure 1, which leads to an authoritative site about U.S. immigration (http://travel.state.gov/visa_services.html) (see Figure 2).

Figure 1. One Result from “Green Card” Search in Yahoo!

Compared to traditional cataloging, what do we observe about the two pages in Figures 1 and 2?

• Traditional cataloging data are absent, nonstandardized, or nonapplicable. There is no reliable information about authorship, title, date, and place of publication.
• A fuzzier idea of discrete work—where are its boundaries, and are boundaries a useful way to think about the content?
• Many more contextual cues about the “spin” or point-of-view of the work conveyed through its organization and graphic packaging (for a thoughtful approach to “spin,” see Crowe, 1986).
• Heterogeneous, unstructured, popular subject vocabulary.
• Commodification of subject terms.
• Cross-referencing through hypertext.
Visa Services

Go to INS to download the new Affidavit of Support (Form I-864) Package

General Information on the Affidavit of Support (Form I-864)

Checklist for the Affidavit of Support (Form I-864)

LASER VISA TO REPLACE BORDER CROSSING CARD

Visitor and Student Visas
Immigrant Visas

Figure 2. U. S. State Department Site.

Subject Access Shortcomings

The "green card" question exemplifies Berman's longstanding critique of much subject cataloging, starting with the problem of applying a nineteenth-century idea of indexing to the twentieth-century scene. The limits of the Dewey Decimal System and the plethora of specialized indexes indicate that the world cannot be organized into a single coherent vocabulary. And even if it could, it can be easily argued that data structure should be different for different disciplines, professions, and populations. Attempts to organize diverse information such as news broadcasts, chemistry literature, visual materials, and fiction in one truly useful vocabulary have been less than successful.

Berman also criticizes the lack of organic holistic cross-references in vocabularies such as LCSH. Extensive synonym relationships and inclusion of popular vocabulary, Berman argues, is a crucial element of any subject vocabulary. It is worth noting that his HCL system of subject headings is far more associative than hierarchical with many more "see also" references than broader-term and narrower-term relationships.

Subject Access Alternatives

Networked information such as that found on the Web offers some alternatives to these dilemmas. This is enabled, most obviously, by the interlinked nature of the Web and relational databases: hypertext is literally a cross-reference (of course, hypertext also enables not-so-useful linking of information, the most prevalent being the lists of links so prevalent on the Web). Richer search results are also enabled by the presence of more searchable content. Simply, there is more data in a given work to search—one is not searching simply a cataloging record (a description of a work) but more of the work itself. This can be used for indexing in new ways, as we see in
search engines based upon prevalence and proximity of terms (for an overview of mechanisms at work in search engines, see Steinberg, 1996).

However, full-text searching and its multimedia equivalents should not be the last word in indexing. Computing presents an opportunity to rigorously interlink diverse vocabularies—to create a thesaurus of thesauri. The Getty Institute’s *a.k.a.* project (http://www.gii.getty.edu/vocabulary/aka.html) attempts to put this idea into action. The *a.k.a.* initiative attempts to cross-reference search terms across several subject authority files. The governing idea is this: it does not matter what you call something as long as it is linked.

Perhaps the most interesting alternative to traditional cataloging presented by networked information is the opportunity for self-determination in indexing and retrieval. The Web presents the clearest evidence of this; self-publishing is the norm, a byproduct of keyword indexing of these works' titles and major subdivisions.

Traditional cataloging rules such as LC's emphasize the assignment of subject terms based on title keywords and the "rule of specificity," but this filtering process has a perverse tendency to remove the author's voice from searching for works by subject, sometimes so much as to render the work invisible. While the intent of subject cataloging is enhanced retrieval through normalization of terms, often this is not the effect (quantifying the effectiveness of subject cataloging is debated by Mann, 1997).

This is not to argue for abandoning subject access in favor of keyword searching, however. Even with sophisticated query languages, keyword searching of full-text databases, citation indexes, and Internet search engines reveal definite shortcomings.

To examine these strengths and weaknesses more closely, we will examine subject access to *Escape Velocity* (Dery, 1996), a nonfiction work, in two contexts.

**Seeking Cyberculture**

The first context is Library of Congress subject access. How might we get to *Escape Velocity*? Searching the Library of Congress catalog (http://lcweb.loc.gov/catalog) by title reveals the results illustrated in Figure 3.

In terms of subject access, we observe the familiar bibliographic standard with its reliable provision of title, author, publisher, date, and physical description. We also find subject access: "Computers and Civilization" and "Internet(computer network)—Social Aspects." The subject headings describe the book's content to a degree—a limited degree.

One of the limits is the number of subject headings assigned, a longstanding criticism by Berman and one that LC has tried to address in recent years. Within the LCSH vocabulary (20th ed., 1997), subject access would be enhanced by the addition of: *Computer sex, Cyborgs, Fantasy games, Internet (Computer network), and Self-organizing systems.*
Another major limitation is the lack of access to subtopics in the book. In other words, what else is the work about? Additional headings could be assigned for body marking, cyberpunk fiction, and perhaps the social effects of the millennium.

A still better approach would be adding the table of contents to the record. A particular advantage to adding table-of-contents information returns us to the issue of self-determination in subject access, as will be seen later in this discussion.

In light of these limitations, we next observe how *Escape Velocity* identifies itself on the Web. Searching for some of the suggested headings above (again in the Yahoo! search engine) leads to the self-promotional site illustrated in Figure 4.
Exploring the site, we see these major supplements: author information, excerpts of (selected) reviews, and a table of contents (see Figure 5).

First, note the evocative words from the table of contents: "synth-rockers," "cyberdelia," and "mechanical spectacle." The words are unsystematized, jargonistic, and perhaps ephemeral, but they express the work in a way that standardized vocabulary would obscure. (Interesting as well that these idiosyncratic terms might escape search-engine stop-word lists, unlike generic terms like "cyberspace," "computer," and "Internet"—the number and nature of stop words in Internet indexing is a rich topic in itself.)

For comparison, consider grassroots cataloging. These descriptors (with underlined terms hyperlinked to a definition) are assigned to Escape Velocity in a site about Cyberpunk authors (http://euro.net/mark-space/bkEscapeVelocity.html): nonfiction, cyberculture, cyberpunk, identity, culture, posthuman, future, Pat Cadigan, William Gibson, Mark Pauline, Stelarc, social history, edge, and 1990s. Besides explicit descriptors, we also observe indirect cues to content. These are conveyed through, first, the URL (the sub-subdirectory of a small commercial site conveys a different impression than a Federal agency, for example). Other cues include editorial style (Cyberpunk-speak), site organization (blending excerpts with press kit), graphic design (techno-chic, tending toward the ominous), and related links (to the presumed milieu of the book).

**SEEKING PROGRESS**

These comparisons are intended to show some cataloging alternatives presented by networked information, perhaps for adoption into
standard cataloging practice. To develop these alternatives further, I con-
clude with recommendations for developers, catalogers, library educators,
and reference librarians.

For developers and catalogers, first, continue to develop subject and
keyword indexing systems; both are useful. Second, enable interlinking
of existing vocabularies through construction of thesauri. Getty's *a.k.a.*
project, described earlier, is an important step in the right direction.
Another experiment to follow is the graphically oriented "hypertextual
searcher's thesaurus" of Johnson and Cochrane (1995). Third, adapt
metadata standards to reap the indexing benefits of traditional and new
media. Specifically, adapt MARC to accommodate new media. The use of
the MARC 856 field for URLs and the use of Web-based catalogs using the
Z39.50 standard are positive steps. Fourth, work toward standards for
metadata (standardized descriptive information embedded into electronic
works). In the short term, seek the integration of metadata into HTML
(HyperText Markup Language) (the HTML "meta" tags move toward this
goal. See http://www.w3.org/TR/WD-html40-970708/struct/global.
html#edef-META) and SGML (Standard Generalized Markup Language).

Metadata initiatives are moving along slowly; track the progress in an
IETF (Internet Engineering Task Force) draft (http://www.ietf.org/
ID.html) and in the report of RLG's January 1997 Metadata Summit
(http://www.rlg.org/meta9707.html).

Finally, to address the issues of transience and fluid boundaries in
hypermedia, work toward development of persistent identifiers for net-
worked information. Current initiatives include OCLC's Persistent Uni-
form Resource Locator (PURL) service (http://purl.oclc.org/OCLC/
PURL/SUMMARY) and the proposal of a Digital Object Identifier (DOI)
(http://www.doi.org) (for a useful evaluation of DOI by Lynch, see http://
/www.arl.org/newsltr/194/identifier.html).

For educators and reference librarians, be conversant in the languages
of different disciplines. When teaching, encourage users of electronic
resources to expect changes of interface, syntax, or dates of coverage.
Concentrate instead on technology-independent methods for seeking and
evaluating information. After all, media are transient; thinking critically
about media content is not.

NOTE

Commodification through sponsoring search terms in popular search engines. When a
searcher enters one of these terms, the sponsor's ad displays as a sidebar. For example,
a search for cars or Phillies or Planned Parenthood could result in links to Honda or ESPN—
or to Operation Rescue. A given query is thus linked, behind the scenes, to a particular
kind of "related term."

REFERENCES


