The information universe: will we have chaos or control?

by Arlene G. Taylor

Technical services librarians organize information by identifying it, acquiring it, providing access to it and locating copies of it. These skills will be important in organizing information on the information superhighway and other online sources.

There is a rumor abroad in the land that goes something like this: In the electronic library of the future we will no longer need technical services librarians.

In November 1992, Newsday carried a long two-part article by Joshua Quittner about the Internet.(1) Those familiar with the Internet know that users can send electronic mail anywhere in the world, gain access to numerous library catalogs and other information banks, download files containing all kinds of materials "published" in electronic form, pursue hobbies, argue politics, or debate any number of topics through "discussion groups" or "meetings." One part of the article particularly drew my attention:

Several problems have limited

the Internet's appeal: much of the

information on it is technical and

scientific; and many of the interactive

uses seem trivial. Most troublesome

of all, it contains so much

information that it's hard to find

what you need, or even figure out

how to look for it.

Within the past three years, however,

new ways of navigating the

networks have been developed,

helping users to scan thousands of

databases by making selections

from central menus.(2)

The article then quoted Ed Krol, author of The Hitchhiker's Guide to the Internet, who observed, "What we had was a library where all the books were dumped on the floor and there was no card catalogue. Now there's a card catalogue and people are starting to put books on the shelves."(3)

I am fascinated that computer experts use library analogies to describe what is going on and what is needed; but at the same time people are saying that the organizing components of libraries are passe. However, Krol's analogy is just that--an analogy. What is really going on is that principles developed by technical services librarians to organize printed information during the last century-and-a-half (if one starts with Panizzi) are being used, almost unwittingly, by people (including trained librarians) to bring order out of the chaos of online information.

I say "unwittingly" because much of the time there does not seem to be a conscious transfer of the knowledge of organizing principles from the world of print to the online environment. People in computing either don't know about the principles of organization developed by libraries in the last 150 years, or they seem to think that they couldn't possibly be applicable for electronic data. So they are reinventing the wheel.

Perhaps part of our difficulty in explaining ourselves to the world lies in the terminology we use. The words "technical services," "cataloging," and "classification" often evoke for nontechnical services librarians a picture of back rooms filled with book trucks where noninteracting people create 3x5 catalog cards, or perhaps enter data into formats on computer screens. For people not trained in librarianship, the image of "technical services" is even less accurate. They think of "technical" as fixing the machines and electrical parts of the library.

Language is tremendously important because it affects everyone's perception of what we are actually doing. But we are now faced with moving into a new era, comparable to the one brought about by the invention of movable type. Just as there were then no words for the concepts of "printing" or "edition," there are no words for many of the concepts we now deal with. Take for example the terms "discussion group," "conference," and "meeting," all used to describe a group of people who all subscribe to the same "list" (another word with a new meaning) so that everyone who subscribes sees everything that everyone else "says." None of these words really expresses the reality that everyone does not "hear" what is "said" at the

The information universe: will we have chaos or control?

same time; instead, people read the same message at different times--hours, days, or even weeks later--and then may respond when the "discussion" has moved on to a different topic (not to mention that several different topics may be discussed at the same time).

Words fail us

So now our words for the organization of print materials are inadequate for many new concepts involved in the online environment. For a few years now a number of us have been using the term "bibliographic control." This term may also be criticized as inadequate because "biblio" is at root a word for "book." However, "bibliographic" does seem to convey the idea of orderly lists of information-bearing entities that are not necessarily books. This definition is using "information" in its widest sense: anything taken into the brain through any of the five senses (e.g., art, music, spoken or printed words). And, that being the case, perhaps it is time to start talking about "information organization."

Information organization may be described as having four aspects:

1. Recording the existence and identity of all types of information-bearing entities, printed or otherwise, as they are produced In libraries this aspect has been most evident in the receipt of publishers' catalogs and announcements. Reference tools such as Books in Print are also products of this aspect of information organization. Now we are getting announcements on the Internet of documents available for downloading from "ftp" (file transfer protocol) sites, with lists of such documents available for each site. I subscribe to an Internet "list" for catalogers called "Autocat" and one for library and information science educators called "Jesse." Through these "discussion groups" I have received numerous announcements of documents available only online.

2. Systematically acquiring these information-bearing entities in libraries, archives, archival Internet communication files, and other depositories. In libraries and archives this aspect has traditionally been carried out by acquisitions librarians and collection development officers, who make decisions about which items will be acquired and stored at a particular location, and who then proceed to acquire those items. As documents have multiplied and budgets have decreased, the decision to acquire an item has included consideration of what was being acquired elsewhere, to avoid needless duplication of expensive or little-used materials.

Until recently, libraries were not responsible for the acquisition and storage of machine-readable information.

This function was the province of computer centers, where computer data tended to be stored as separate entities (e.g., magnetic tape or disks). Increasingly, files of information are being stored online and libraries are getting involved in acquiring the right to make them available to their clientele without actually acquiring anything physical. For example, many libraries no longer subscribe to the print version of some indexes, such as the New York Times Index; instead they purchase the right to have their users search the online version. At some level in computer networks, someone has to decide which information packages to make accessible online to everyone and which to place in archives. In the latter category it is necessary to decide which items are worth keeping for posterity. Will some be discarded? (e.g., are messages sent through online discussion groups worth keeping like internal memos at the White House, or are they disposable notes posted on office bulletin boards?) Will some files be placed in archives? If so, where?

3. Providing name, title, subject, and other useful access to information-bearing entities. Library catalogs (most with authority control) and periodical indexes (most without authority control) are the prime examples of provision of access to documents through name, title, and subject. In addition, bibliographies and other reference tools can provide such access. The Internet includes lists of documents available at many of the sites where a group of data files are archived, and most of these can be searched by keyword; but so far there is nothing as sophisticated as a catalog or index. At this time we are talking about no more than a few hundred documents at each site; so keyword access is more or less satisfactory. But what about the day when there are thousands of documents at each site? And how will we know which site has which documents?

Many people seem not to realize that name, title, and subject access in any of the tools just mentioned must be provided by people--people who have a thorough grounding in the principles of information organization. Pr been written to try to provide such access automatically, but without complete success so far. One of the most successful is CARL's UNCOVER, an automatic indexing of the tables of contents of serials. But not all periodicals and serials have a table-of-contents page. And the system makes errors; in the field of architecture, for example, UNCOVER sometimes mistakes the name of the architect of a structure that an article is about for the author of the article. Catalog records, entries for most bibliographies and indexes, and any authority control that is provided are still created originally by people. Libraries have become much more accepting of externally produced records, so copy cataloging has become somewhat of a "check-in" operation; but someone must create the externally

The information universe: will we have chaos or control?

produced records in the first place. Authority linkage is a time-consuming and heavily human, albeit much appreciated, activity. Subject analysis so far has defied automatic techniques such as word counting; only a human can attach words to a concept that is the subject of a document but is never explicitly named in that document.

Putting the right data into records and providing for authority linkage are only part of the process of providing access. The other part is designing systems for searching and display that will allow information-seekers to find easily what they need and want. Our current information-organization systems vary widely in this respect. Many of today's online catalogs still do not even provide cross references for subjects; meanwhile, the rest of the information world is moving on to create catalogs providing such advanced search and retrieval methods as "closest-match" retrieval and ranked retrieval based on feedback from the searcher who says in essence, "Give me more like this one."

4. Locating copies of the information packages. Traditionally, the prime way to know where to locate particular items has been to search catalogs that show holdings of libraries and other collections. Lately, circulation data have been added to show the actual presence or absence of an item at a location. Union catalogs such as the OCLC and RLIN databases allow us to see which locations physically own particular items. Many library catalogs are available on the Internet, letting users learn not only which libraries own an item and whether it is checked out at that location, but which have it on order and when it is expected to arrive.

Indexes and bibliographies in the past have not helped users learn where to find copies of items identified in those sources. However, as resources have become more scarce it has become increasingly important to be able to locate copies of analytical items that are held locally, such as articles in periodicals or papers in conference proceedings; so librarians have begun mounting indexes in databases alongside catalogs, creating some bluffing of the distinctions between catalogs and indexes.

As more users become aware of online documents available through the Internet, information-organization tools will need to provide lists of Internet addresses for various "ftp" sites so that documents can be downloaded from them. I mentioned earlier that I have received announcements of documents on Autocat and Jesse; my finding these documents is greatly impeded if I forget to print the ftp address before I delete the message announcing the document's availability, since there is no union catalog of online documents.

After copies of a document are located, the next step is to get them into the hands of users. Traditional interlibrary loan methods have been slow and labor-intensive. But with the availability of location information on the Internet, we are moving closer to the time when a user can submit an electronic request for the delivery of an item that, depending on its nature, may then be sent via electronic means or via traditional ILL "snail mail." Commercial ventures already provide documents for a fee in response to electronic requests.

A desperate need

So it seems apparent that the online world is desperately in need of the skills we librarians have developed in the area of information organization: making new information-bearing entities known; acquiring such entities at certain points of accumulation; providing name, title, and subject access to these entities; and providing for the "physical" location of copies. These functions are in addition to the continuing need to deal with print and audiovisual materials. Recent decades have seen an exponential rise in the number of publications, with no sign that this will end soon.

Michael Gorman has asserted that there should be two core elements in a library science curriculum.(4) The first is the history, philosophy, and ethics of the field; the second is bibliographic control, because, Gorman points out, bibliographic control (which I am calling "information organization") is the way we, as librarians, think and should think. It is the essence of what we do and how we do it. That is, learning itself is based upon the ability to analyze and organize data, information, and knowledge.

In the area of reference Gorman observes that reference librarians cannot do their jobs without this ability: They must think in an organized way, classify and catalog the question, and then match the category in which the question has been placed to maps of knowledge and information either in their heads or in some information source they identify that will provide such maps.(5) In addition, the bibliographies, catalogs, indexes, and other materials that are the standard equipment of a reference librarian are the products of the world of information organization; to use them, one must understand their structures. Lois Chan wrote that consulting a reference librarian who does not understand the catalog is like consulting a doctor who does not understand the nature and function of drugs, having never studied pharmacology in medical school.(6)

The future will see reference workers, available to users by telephone or electronic mail, who understand the structures of the Internet and other online information

The information universe: will we have chaos or control?

sources in order to help people find what they want. Such people will function in much the same way that people are now available at 800 numbers to help with the installation and running of various computer programs such as WordPerfect. Helping people through the maze of online information sources such as the Internet is very similar to helping people through a maze of reference books, catalogs, and classification schemes.

Another example Gorman uses is the area of collection development. Collection development librarians also cannot work without information organization. They use bibliographies and catalogs, and they rely on the details recorded in catalog records more than almost anyone else. Much of their work requires the use of a classification scheme; for example, approval and other gathering plans are based upon classification schemes.(7) Collecting online documents at various Internet sites requires similar skills.

Administrators, too, must know about information organization in order to make informed decisions about technology. They must be aware that organization of information is the only alternative to electronic chaos. Cutting costs by trying to cut back on the information organization process will only result in putting libraries at the margins of the information-access field.

Is it really possible that with everything stored in the computer, technical services librarians are not going to be needed much longer? That will be the case only if we do not quickly adapt our skills to the increasing need for organization of information in all parts of the information universe. An exploration of the world of information organization and what we do in it shows that not only is there still a need for persons to create the organization of information, but there is also a desperate need for people who are thoroughly grounded in the principles of information organization to interpret and manage the world of information and to design the enhanced and expanded access systems that will be needed to exploit the information universe.

We probably should also adapt our terminology, which affects not only how others view us, but how we view ourselves. We are, and always have been, information organization specialists. Such terminology will assist those outside our specialty to understand our range of expertise much better than use of the words "technical services," "cataloging," or "bibliographic." We may also need to give up the word "library," but that is the subject of another essay.

Notes

(1.) Joshua Quittner, "Plugged in: Internet plays growing role as world's electronic highway," Newsday, Nov. 2, 1992, pp. 3, 29; "Getting up to speed on the computer highway: Overcoming real problems in a virtual world," Newsday, Nov. 3,1992, pp. 51,54-56. (2.) Quittner, "Getting up to speed," p. 51. (3.) Ibid. (4) Michael Gorman, "How cataloging and classification should be taught," American Libraries 23, no. 8 (Sept. 1992): 694-697. (5.) Gorman, p. 696. (6.) Lois Mai Chan, article in "Catalogs and catalogers: Evolution through revolution," edited by Tillie Krieger, Journal of Academic Librarianship 2, no. 4 (Sept. 1976): 176-177. (7.) Gorman, p. 696.

ARLENE G. TAYLOR is associate professor at the University of Pittsburgh School of Library and Information Science.