Special Section: Happy Birthday to MELVYL® (Part 3)

The Next Generation of Public Access Information Retrieval Systems for Research Libraries: Lessons from Ten Years of the MELVYL System

Clifford A. Lynch

This paper views the design of the next generation of public access information retrieval (IR) systems in higher education from the perspective of a decade of development, deployment, and operation of the MELVYL online system at the University of California (UC). It highlights design decisions and assumptions that were made for the MELVYL system that have proved advantageous, as well as those that have proved limiting or have led to dead ends. Our design choices were probably similar to those made by most other online catalog designers at the time. Some decisions at UC that have proved in hindsight to be shortsighted or cowardly (and also a few that proved better than we might have hoped) were only guesswork, because there was no base of experience from which to work. Other decisions were artifacts of limited functionality and capability from the underlying base of information technology upon which the catalog was built, or of a limited budget to acquire resources. Particularly in the case of computing hardware, it was not that desired technology did not exist ten years ago (unlike certain supercomputing applications—visualization being the most striking example—that emerged during the 1980s), but that the cost of the desired computing cycles, memory, and mass storage was out of reach. Costs of these resources have dropped now sufficiently that they can be used more freely as we consider systems for the 1990s.

The available base of software technology was a different matter. The limited functionality in the software components, such as database management systems (DBMS), that might be used to build an online catalog was a serious problem. In 1980, the DBMS choices were few, and none of them was entirely satisfactory. Interestingly, as we consider future directions for the MELVYL system in 1992, the choices seem to have improved little in terms of functionality, although the available commercial software has matured considerably in terms of stability and performance. The full set of functionality still seems tantalizingly out of reach, manifested most broadly in database systems that remain as research vehicles within the computer science research community, and thus unsuitable for production use in a system the scale of the MELVYL catalog.

Finally, in terms of delivery platforms, we viewed the system as limited by the installed base of character mode ASCII terminals and so designed to the lowest common denominator "glass teletype." In theory, we might have procured a special terminal for use with the MELVYL system (as some other systems had done), since, as discussed in more detail later, our initial assumption was that most terminals for catalog access would be placed in libraries. But we felt it was important to be able to support the installed base, presuming that networking on the UC campuses would continue to improve and that over time more of this installed base would be able to reach the catalog. Given the explosion of networking that occurred later in the 1980s, this proved to be a very wise decision as it greatly facilitated wide access to the catalog.

Clifford A. Lynch is Director, Division of Library Automation, Office of the President, University of California.

MELVYL is the registered trademark of The Regents of the University of California.

The history and current status of the MELVYL system has been amply covered in the papers that have appeared in the two previous "MELVYL at Ten" special sections of *Information Technology and Libraries* and in the spring 1992 issue of the *DLA Bulletin*. But a review of the design assumptions and system objectives for the original MELVYL online catalog, many of which, to my knowledge, were never explicitly articulated and debated as part of the planning process prior to its development, forms an essential part of the context for this paper. Thus, the first part of the paper reviews them with the benefit of ten years of hindsight, along with certain realities of the information technology base of the late 1970s. The remainder of the paper focuses on key problems that emerged as we gained experience with patron use of online catalogs at UC and elsewhere, and as the MELVYL system has grown larger, more complex, and more capable. As we consider requirements for future systems and the future evolution of the MELVYL system itself, I believe that we must revisit some of the basic assumptions. Thus the paper concludes with a discussion of some of the new possibilities opened up by altering the fundamental assumptions that guided the development of the MELVYL catalog in 1980.

DESIGN ASSUMPTIONS AND OBJECTIVES OF THE ORIGINAL MELVYL ONLINE CATALOG

The first and most important thing to recognize about the original design assumptions underlying the MELVYL catalog of the early 1980s was that it was designed to be an online catalog. It was not envisioned as a more general information access system. There was no discussion of including abstracting and indexing (A & I) databases (beyond speculation that this might be a desirable expansion in some distant future), full text, or images, and there was no thought of having the system serve as a gateway to a wide range of information resources (both academic and commercial) available through the network. At the time, there was no real national, much less international, network that would have made such a gateway function possible even if it had been deemed desirable. In fact, when the MELVYL catalog appeared on the DARPA Internet in the mid-1980s, it was one of the first online catalogs to be publicly accessible through the Internet and thus helped establish the precedent that led to the availability of a rich collection of public access Internet-accessible resources today.

When the MELVYL catalog was being planned, there were few examples of operational online catalogs to guide the development, and those that existed were both so new and so poorly instrumented that little information could be gathered about how library patrons used these new tools, other than that most users of online catalogs seemed fairly enthusiastic about them.[1] The MELVYL catalog was conceived of as a conservative, straightforward mechanization of an existing physical card catalog, since the traditional card catalog was a well-proven, well-understood access tool, despite its limitations.

The functions designed into the MELVYL online catalog closely mirror those found in traditional card catalogs: searching by author, subject, and title. Keyword access, as an alternative and supplement to exact searching of author, title, and subject fields, was perhaps the major innovation visible to the user. In the early days, there was substantial controversy about whether it was worthwhile to offer subject searching, since the conventional wisdom of the time suggested that users seldom did subject searching, even though it could be accomplished through the card catalog.[2] This misconception about online catalogs was resolved by a massive, two-part study funded by the Council on Library Resources, *Users Look at Online Catalogs*, that appeared in 1982 and 1983 and demonstrated that subject searching was a major search mode for online catalogs.[3,4]

Many of the later enhancements, such as access by language of publication or date of publication, were controversial because they did not mirror capabilities of traditional card catalogs, and, indeed, could not be assessed within the experience of a card catalog user. Frankly, I believe that many of them became part of the system more because no one provided a compelling reason not to implement them than because they were considered a particularly exciting idea by the community at large. Most of these extensions have been at least modestly successful and invaluable for certain specialized situations, such as research in the early history of publishing.

To understand fully what is, in hindsight, a conservative functional definition of the on-

line catalog, one must recognize that the political and organizational commitment that led to the development of the MELVYL system was much more complex than a simple desire to improve access to a given library collection by developing a computer-based access tool that was more effective than the traditional card or book catalog. Certainly this was identified as a desirable goal, as were many of the other benefits that are typically attributed to online catalogs (elimination of card filing and the space taken up by the card catalog, no more worries about patrons tearing cards out of the catalog, availability of more current information to the patron, etc.). But the plan for the MELVYL system emerged out of a document authored by then assistant vice-president for Library Plans and Policies Steve Salmon, The University of California Libraries: A Plan for Development 1978–1988,[5] which looked much more broadly at the strategies for developing and funding the roughly one hundred libraries of the nine campuses of the University of California as a unified system for the first time. This plan called for initiatives such as the establishment of regional storage facilities and enhanced intercampus library materials delivery systems.

The key role of the MELVYL catalog was to provide a library user at the University of California with a coherent view of the collections held throughout the UC system as a whole, from any library within that system. This was the compelling justification for developing the MELVYL online union catalog, rather than simply suggesting that individual campuses develop online catalogs at their own pace, as their own means and priorities permitted, to improve patron access to their local collections. Local, campus-based online catalogs had great appeal because they maximized local control. One idea discussed during the initial planning of the MELVYL catalog was a distributed union catalog of nine linked campus online catalogs that would have addressed the desire for both local control and a union catalog. This idea was rejected—correctly in my assessment—as too difficult and risky, from both technical and management perspectives.

In my opinion, a union catalog built out of a set of linked local catalogs would also have been compromised by the natural desire of each campus to emphasize local needs, including requirements for technical processing, as opposed to public access; university-wide public access would have been a lower daily priority for campus library management. Thus, not only would a distributed union catalog have faced formidable technical and operational management problems, but it probably would never have obtained the necessary management focus and commitment to develop and mature. By assigning the responsibility for developing the union catalog to a separate organization (what became the Division of Library Automation at the Office of the President, the "corporate headquarters" of the nine-campus system) and limiting the scope of the MELVYL system to public access rather than including technical service support, a strong and focused development effort toward an online union catalog was ensured.

The policy commitment to employ computing and telecommunications technologies to provide unified access to the collections of the university within the context of the overall plan for developing the UC libraries as a system had obvious major benefits. We received great support both within the university and from the state legislature, which provided funding to implement the Plan for Development as a whole, including development of the union online catalog. The arguments for offering universitywide collection access were more compelling than simply the claim that converting to an online catalog would somehow make the libraries' services "better," particularly at that time and given the lack of experience with online catalogs at other institutions.

Having decided to build an online union catalog, we proposed (or more precisely, simply decided, without much debate) a rather direct and conservative approach to automation of the card catalog as most viable. It minimized controversy and technological risks, particularly given the lack of data and experience then available to support and guide a more ambitious and adventurous design. The enormous scale of the MELVYL catalog (originally planned to support about one thousand terminals and ultimately to hold at least six or seven million unique titles representing perhaps ten to fifteen million holdings), particularly compared to systems in other libraries being planned at the time, also argued strongly against a highly experimental design, either in terms of function or underlying computing technology.

It was essential that the design for the MELVYL system be understandable and achievable. Great care was taken to distinguish the MELVYL project from more visionary (and, in implementation, more illusive) projects ranging from Bush's MEMEX to Ted Nelson's visions of worldwide hypertext, and to avoid discussing it in the context of revolutions in information access. The MELVYL system was intended to be a production online union catalog, not an experimental attempt to construct the electronic library or information utility of the future. UC policy has never recognized the objective of building an electronic library; the view has been that information technology should be applied aggressively but judiciously both to expand the scope of the UC library collections and access to them within the intellectual framework of existing library missions and service objectives.

Given the historical context of the commitment to build the MELVYL catalog, it is interesting that as of 1992 the majority of the UC campuses have either developed or purchased campus-based online catalogs. There are a number of reasons for this. One is that given the critical role of the online catalog in library operations and services, there is a natural desire for campuses to want control over their own catalogs. Second, there are operational advantages available to a library from an integrated system; all of the campus catalogs are part of integrated library automation systems. In the mid-1980s, some people, including myself, hoped that developments in computer networking protocols for library automation would enable us to integrate the system by linking the MELVYL online catalog and campus systems to support cataloging, database maintenance, circulation, and other functions.[6] Progress in this area was much harder and slower than expected, and the library automation vendors resisted doing work in this area, seeing little market advantage.[7] In 1992–1993 we expect to attempt the first implementation of such a link between the MELVYL system and a campus system: A Data Research Associates (DRA) system at UC Davis will support circulation, and the MELVYL system will pull circulation status for Davis material out of the DRA system in real time for display to MELVYL users. A third factor in the development of campus catalogs was a crisis of confidence in the long-term viability of the MELVYL system. A series of policy, management, and technical problems related to the scaling up and long-term support of the MELVYL system in the mid-1980s led, I believe, several campuses to want an alternative catalog available that they could control directly.

Over time, it has become clearer that the MELVYL system and the campus catalogs fill complementary but increasingly distinct roles, and that there is great value in having one to back up the other for the limited set of services that are duplicated. As intersystem linking technology improves, the campus systems and the MELVYL system will grow even more complementary, and the growing set of options (as illustrated by the effort with UC Davis, assuming that it is successful) will offer the UC campuses expanded choices in developing strategies for campus library automation.

IMPLICATIONS OF MECHANIZING A CARD CATALOG

From the basic assumption that the MELVYL catalog would largely mirror the functions of a traditional card catalog, several design principles fell into place with little critical examination.

The Patron Would Go to the Library to Use the System

This principle pervaded all aspects of the system. Terminal installations were planned for the libraries, with some discussion of the desirability of eventually locating a few terminals in other key locations such as dormitory lobbies. The user would go to a MELVYL terminal in the library (or, a faculty member, fortunate enough to own a terminal, could dial up the system using a modem) and perform searches, locate material, and then go to the library to use or borrow it.

This view of the world, which now seems charmingly antiquated, makes sense when one realizes that in 1980 personal computers did not exist, and local and wide area networks were essentially experimental curiosities used by the computer science research community. Today, we view online catalogs in the context of a richly networked distributed computing environment where we increasingly assume that all users have ready access to the network and have considerable desktop computing power. In this environment, it makes sense to consider electronic links to document deliv-

ery service, electronic delivery of documents (remember that low-cost, high-quality laser printers, now ubiquitous, also did not exist in 1980), and various forms of system-initiated current awareness services. In an environment in which the majority of the user community had access to the system only by going to the library, such services were impractical and did not make sense within the existing information technology infrastructure.[8]

(A careful look at the history of the MELVYL system's evolution might indicate a broader commitment to remote access on the part of the system developers than was politically admissible at the time. We did not really believe that patrons would always have to visit the library to use the online catalog. We designed for a lowest common denominator terminal to make the system usable from many locations. We moved far more aggressively than most online catalog projects in trying to make the MELVYL catalog accessible through campus networks, national networks, and even campus port selectors as opportunities arose during the 1980s. However, it is only now, in the 1990s, that we are seriously working to provide facilities such as the ability to send search results through electronic mail, electronic mail-based current-awareness services, and electronic links to campus document delivery services, as well as seriously planning for network-based document delivery. We were not always forthright about the extremely high priority we gave to network accessibility of the catalog, and in the early years of the development of the MELVYL system we probably devoted more effort and resources to this goal than the policy-makers would have been comfortable with had we showcased the effort.)

In summary, this original design principle was realistic when it was adopted but eventually became totally wrong as the information technology infrastructure at UC developed. Fortunately, we were cautious about taking it too seriously, and fairly quick to abandon the plan as its focus became outdated.

The User Would Always Remain in Full Control and Be Fully Aware of the System's Function

There was considerable and well-justified concern about designing a system that would try to do too much for the user and that, in the end, would either fail to deliver desired results, particularly to users who knew what they were doing, or would deliver results that were not comprehensible to users, particularly to the great majority of users who lacked—and did not desire—intimate knowledge of the system's inner workings. Too many people, in 1980, had seen the results of inept system design, particularly in contexts where computer scientists or information retrieval researchers had built prototype library automation systems that were either unusable, incomprehensible, or both. The experimental IR systems of the 1970s, such as Salton's SMART system, which used complex automatic indexing and query evaluation methods, were superb examples. Mike Berger accurately characterizes this type of IR system, from the user's perspective, as magic. There was a strong consensus that the patron must understand what the system was doing and remain in control of interactions with it. There was, perhaps, a bit of wishful thinking embodied in this principle. Few users of either traditional card catalogs or online catalogs really "know what they are doing" (except trained librarians). Results of a subject search against a large database cataloged with Library of Congress subject headings are certainly neither obvious nor intuitive to most library users, I believe, and studies of the difficulties of subject searching seem to support this contention.

These considerations led to the design of a system with two modes. One was a rather simple menu mode, called Lookup mode. As the system developed, the Lookup mode was kept simple and was an increasingly limited subset of the overall capabilities of the MELVYL system. It was not even implemented for the A & I databases when they became part of the MELVYL system. Lookup mode was probably a necessary evil given that in 1981 the MELVYL system was the first interaction that many patrons had with computers, and they seemed more comfortable with a menu-based approach. As computer literacy has increased, users seem more and more willing to use command languages, and as of late 1992 Lookup mode is being discontinued.

The other MELVYL interface mode is a command language. This is a reasonably simple structured command language with full Boolean capabilities that supports queries such as FIND SUBJECT INFORMATION

RETRIEVAL AND TITLE AUTOMATIC with the usual abilities to abbreviate words. Natural language approaches were rejected because we did not feel we could interpret natural language queries in an unambiguous and comprehensible fashion.

Performance issues, combined with the growing problem of users confronted with very large retrieval results as the database grew, led to increased precision in query processing. Keyword searching in titles and subjects was supplemented with exact searching (left anchored with optional truncation). In the early days, subject keyword searches actually searched both subject and title fields of bibliographic records in an attempt to help users with cataloging vocabulary problems. This process was abandoned because of the extra cost it added to the subject searches, because the results were hard for the user to understand, and because it led to even larger result sets. A number of other indexing and access decisions were similarly revisited.

When periodicals were added to the database, they were placed in a separate file, both to help precision and to help system performance. Similarly, when A & I databases were mounted they became a series of separate files, both for reasons of precision and comprehensibility, and because the A & I vendors insisted that the identity of their files be preserved. From the perspective of someone who understands the library, all of these choices do help keep the user in control and the system's behavior comprehensible.

The MELVYL system today, viewed with some detachment, requires modest user library literacy. To really exploit the system, the user needs to understand the difference between titles and subjects, personal and corporate authors, books and periodicals, and periodicals and articles within these periodicals. These concepts seem basic enough to those of us who are seriously involved with libraries; but there is evidence that such knowledge is hardly universal, even within a major research environment like the University of California.

There were other implications of this design principle of total user control and awareness. The original design did not enable the catalog to make assumptions about what the user wanted or intended when issuing commands. It avoided heuristics, spelling correction, or making suggestions that might be helpful in many, but not all, cases. In general, the system simply received orders from the user; it did not try to be overly "helpful." As the system matured, the position on this softened somewhat, largely in response to extensive analysis of transaction logs and other user studies. A few heuristics were introduced, very carefully, and with much controversy. In some cases, librarians at the campuses remained sufficiently distrustful of these heuristics that in bibliographic instruction they trained users to avoid triggering them.[9] However, I believe that these heuristics are helpful, particularly to users who access the system occasionally and casually across the network and are unlikely ever to bother learning how to use the system more effectively.

One corollary of this assumption has been the ambiguous attitude toward instruction in using the catalog. We want a system that can be used without instruction but that repays an investment in learning (through formal instruction, reading the manual, reading the help screens, or, in the future, spending time in a tutorial mode in the user interface) by providing more powerful, precise, and effective facilities. In some sense, these two goals are conflicting, and balancing them is difficult. The staff at the UC libraries has done a superb job of developing bibliographic instruction to support the use of the MELVYL system, and library patrons who invest in such instruction benefit substantially. At the same time, however, there is a natural tendency to focus on the relatively sophisticated and serious user, perhaps sometimes to the detriment of a user who might feel well served by a more heuristic, actively helpful catalog. The counter-argument is that by implementing such features, we are permitting the naive user to shortchange him or herself, and that the system should not allow that user to go away remaining ignorant and satisfied. Realistically, there is a certain percentage of users who are fairly determined to remain ignorant or who use the system too infrequently to bother learning much about it. Thus it seems we may as well make them happy too. The proper balance between these positions is controversial.

I believe that this design decision was both right and wrong; it was right in the sense that online catalogs must be able to present a user interface that meets these criteria fully. But I believe it is best viewed as a requirement, and not a limitation on other functions that may

be offered by the system, particularly as the online catalog becomes a much larger, richer, and more complex information access system. Later sections of this paper will return to this issue.

I think it is also important here to separate the design objective from the system's technical limitations. Supporting very large bibliographic databases is still a major performance problem, which is why the system segments databases such as MEDLINE® into a series of backfiles. It is a major reason why the MELVYL system has not merged monographic and periodical records into a common database. As we design future systems, however, we should not use the design objectives of precision, user control, and comprehensibility to justify the need to recognize technical limitations in the system design. We should recognize technical limitations for what they are. Good response time is a very important design objective, and it seems likely that we will continue to subordinate a number of other objectives in the name of response time.

The Catalog Should Not Provide an Evaluative Viewof the Collection

Early in the design of the system, the conclusion was reached that results would be presented in standard main entry order. It is unclear to me just how this was decided, but even today it is a basic assumption. There was some discussion about adding options in the future to permit users to alter the arrangement of result sets (for example, to ask for a display by publication date), but this function has never been implemented due to technical problems and the cost of required computing resources. It is important to recognize that, as with many issues involved in the design of the catalog, the selection of the default function, service, or option is really the critical choice. Relatively few users know how to override the defaults or choose to do so. Most of these users are fairly sophisticated, for whom the choice of default is relatively unimportant, because they understand that there is a default, what the default is, what other options are available, and they can make a choice. We chose to default to presenting all publication dates and languages, and all works, on an equal basis.

Part of this decision was based on the earlier requirement that the operation of the catalog be entirely comprehensible to the user. This principle argued strongly against various heuristic or probabilistic ranking approaches that might have been applied, either based on the estimated closeness of match between the user's query and the citations in the result set, or by making assumptions about the overall nature of material that would be of interest to the user (e.g., English language and recent publication date for at least a large number—perhaps the majority—of users, and particularly of relatively unsophisticated users). Assumptions about the nature of material of interest to the user also raised questions in the context of the growing sensitivity to cultural diversity issues within UC.

Experience over the past decade has indicated that one of the greatest problems users face is managing the size of retrieved result sets. Searches with zero results are quite common and occur for many reasons (typos, spelling, problems with the subject vocabularies used in cataloging material). The sort of heuristics discussed above help resolve these situations. But even more serious are the more and more frequent large result sets containing hundreds or even thousands of records. Users desperately need help in navigating through these huge results and in reducing their size. In fact, users want much more than assistance in ranking based on known or assumed user preferences and closeness of match to the user's query. They want help from the system in finding "a few good references on . . ." rather than everything written on the topic. This is anathema to the principle of nonevalulation that has guided the development of the MELVYL system. It is also a very complex need to address, since the information available to evaluate material in traditional cataloging records is limited. To meet this need, bibliographies, book reviews, pathfinders, review articles, citation indexes, statistical impact factors, databases defining core literature in various disciplines, and other resources must be integrated with the catalog to produce a complex knowledge base that goes far beyond the contents of existing traditional bibliographic databases. This is not simply a problem in the design of online catalogs; yet I believe it is one of the most critical needs to be addressed in the academic information access systems of the 1990s and beyond.

The User Will Be Anonymous

Users of card catalogs are anonymous. Of course, the card catalog is a totally passive device; it has no memory of its use. So while a library might physically control access to its catalog, there is no need to keep track of who is using it because there is absolutely no benefit to be gained from such tracking. In the design of the MELVYL system users are anonymous. There was no reason to do otherwise; but another part of the consideration was the enormous administrative burden of issuing accounts on the catalog, either through a direct registration process or by building on the registration processes already in place at the campus libraries. Later in the development of the system, as licensed A & I databases were mounted that are restricted to the UC community, facilities were added that allow us to determine if a user is a member of the UC community without having to identify individual users.

From a policy perspective, I believe that users deserve the reasonable expectation of privacy in their use of the catalog in the same sense that they reasonably expect their circulation records to be confidential. If they use the system other than anonymously, they should expect that their searches remain confidential. Certainly, anonymity is an excellent guarantee of confidentiality. But from the point of view of designing an effective information retrieval system, anonymity of users is a great constraint. It precludes viewing user-system interaction from any perspective broader than a single session. The system cannot remember the user's preferences (for example, the user does not read anything but English and French), cannot tell the user about news of interest, cannot support current awareness, cannot remember that the user has not used the system in six months and might like a few reminders or that the user has used the system twice a day for the last year and is familiar with its features. I believe that part of the future evolution of the MELVYL system and other online catalogs will be to include features that are sufficiently valuable that the user will often be prepared to trade the near-absolute guarantee of confidentiality offered by truly anonymous access for non-anonymous access with trust in policies promising confidentiality. Further, there are several design alternatives that could allow the system to remember information about users from one session to another without the need to identify them, such as self-registry with a nickname. These features limit function in that if the system is to interact dynamically with a user's workstation on the network it must know the identity of that workstation, thus compromising anonymity; but they do allow the system to remember user preferences and activity profiles, thus allowing some improvements.

Further, it is possible to develop an infrastructure of trusted "brokers" that can serve as intermediaries to conceal the identity of client workstations in a network environment, if necessary. Such arrangements are already being developed for electronic mail "personals." I feel that future systems should support anonymous access up to the limits required by license agreements. For the casual—or very paranoid—user, this is useful. Any type of registry is a barrier to access for the casual user. But I believe that basing the system on the assumption that the time horizon for user-system interaction is a single search session, we have disastrously limited our ability to build effective information services.

NEW FUNCTIONAL CONSIDERATIONS FOR THE 1990s

Since the deployment of the MELVYL catalog in 1981, two major considerations have emerged that were not addressed in the initial design. Both of these are products of changing technology. The first was the expansion of the system from a simple online catalog to a collection of information resources that included A & I databases, such as MEDLINE and CURRENT CONTENTS®, and gateway access to other online catalogs, A & I and full-text databases mounted on other systems accessible through the network, and specialized resources such as scientific databanks and weather information. The inclusion of such databases was a massive, but in many ways straightforward, extension of the original design of the MELVYL catalog. The command language was extended and generalized, but a user, having learned to use one database, generally has little difficulty transferring knowledge to others. (Some of the databases, such as MEDLINE, have extensive, unique, specialized features; but they are typically functions of the information in the database, not of the user interface.) It was really not until the mid-1980s that the decreased cost of com-

puting cycles and disk storage made implementation of such huge databases feasible, and only continued reductions in these costs have allowed us to mount large numbers of them for intensive use by the UC community. The MELVYL system currently services 500,000 queries a week during busy times of the year. The computing resources to support this would have been out of reach a few years ago.

While the mechanics of using a given database are less and less of a problem, the multiplicity of choices is becoming confusing and overwhelming.[10] It is not always clear to most users when it is appropriate to use a given database or to choose one over another. Further, the fragmentation of information into a large number of databases (typically licensed from commercial A & I services), each with different and idiosyncratic data elements, varying chronological coverage, varying completeness of coverage, and overlapping with other databases in complex, difficult-to-define ways, is immensely troublesome for most information seekers. Other database-specific attributes add to the confusion: full text (for some or all material in the database), images, abstracts, data quality, and timeliness. The number of resources continues to multiply, and the systems of the 1990s will have to help users select appropriate resources for various information needs and effectively combine results from multiple databases. Here again I suspect that the ability to remember the user's preferences and past experience with the system will be crucial. One can, for example, imagine a system that conducts a partial reference interview with a new user and then uses that data to guide the user among available information resources. In the past, such approaches have been impractical because the user would not be willing to invest time in a lengthy background dialogue with the system at every new session.

Just as with the issue of evaluative retrieval and ranking from bibliographic databases discussed earlier, the problem of selection of appropriate resources is not simply solved by adding functionality to a retrieval system. New descriptive directory databases will have to be developed to support these new navigation and selection functions, and the appropriate data elements and descriptive approaches needed to create the information base to support these new functions is still very much a subject of active research.[11] Further, effective user guidance in selection among resources is again an evaluative function.

The second novel design issue concerns the relationship between the information access system and the overall distributed computing and workstation environment that is currently evolving at major academic institutions. One aspect is the development of user interfaces that effectively exploit bitmapped display devices (perhaps with color and sound capabilities), which are rapidly replacing the older, character-oriented terminals for which the current MELVYL system is designed. I do not regard this change as simply the development of graphical interfaces with windows, icons, and a mouse or other pointing device. I believe that harnessing the potential of bitmapped displays to provide effective and intuitive user interfaces to information retrieval systems will be a much more complex and subtle undertaking and will require extensive research, prototyping, and evaluation in the next few years.

Character terminals will persist for at least another ten years alongside the bitmapped displays. We also must consider how to maintain sufficient consistency between the character and bitmapped interfaces so a user can easily make the transition from one to the other, or even use one at home and the other in the office. Driving a bitmapped display will require considerable bandwidth, and a user with a workstation at home may prefer to emulate a character-mode terminal if constrained by a slow dial-up phone line.

There are also software architecture questions to be answered. It is clear that we are moving into a client-server environment, perhaps with multiple levels of clients and servers. But the distribution of function remains unclear. Where will software reside to perform various functions (searching, display, integration of information from multiple sources)—on the central institutional system or the user's workstation? What organization will write, maintain, and support this software? Will the institution or commercial vendors provide software that is installed and supported on each end-user workstation within the institution, and undertake all of the software management and maintenance issues that arise in such a large, poorly controlled, heterogeneous environment? Will workstation vendors provide it as part of the base system, or will library automation

suppliers or third-party software suppliers outside the library automation tradition supply this software? Or will the model for the next generation of software be a desktop device working as a graphical display, driven from centrally administered institutional systems (either purchased or locally developed) using protocols such as the X-Window system? (In this last scenario, these institutional "user interface servers" offering X services to end user clients might themselves function as Z39.50 clients to institutional or national information servers on the network.)

A closely related question is the degree of centralization of access to information resources. Will the end user directly access a multiplicity of resources from his or her workstation (and even, in many cases, pay the costs of using them, and negotiate license agreements, in which case software on the workstation will have to support mediating and integrating functions)? Or will most access to information resources be institutionally funded and mediated through institutional systems? A full exploration of these issues in the context of the changing library and the developing world of networked information goes far beyond the scope of this paper[12], but it is important to recognize the implications of these policy issues for the architecture of future systems.

It is also important to recognize the immaturity, both in conceptualization and technology, of the distributed computing environment as a context for information access systems. This may be at about the same point as network access to online catalogs was in 1980. We know it is important, but the details are not clear yet. The long-term implications, both technical and organizational, are unknown, and developers of production systems can only track the evolving environment closely and be prepared to continue to adapt the system to its requirements aggressively.

CONCLUSIONS

The evolutionary descendants of the MELVYL system (and other systems of its generation) will differ in many ways from today's online catalogs. Certainly, there will be changes which exploit better delivery technologies for the user interface, such as bitmapped display devices. But, in a more profound way, I think that they will not be designed as mechanizations of the old card catalogs, but as new information access and delivery platforms which operate on databases that are much richer and more complex than today's bibliographic and A & I databases, and that are built by combining and integrating information from multiple sources. Not only will the user interface and searching algorithms change, but the contents and scope of the information bases to which the system provides access will change also. These future systems will be more heuristic, and will evaluate information and guide the naive user, while still permitting the "expert" user total and direct control. They will be more generous in their expenditure of computing resources to help the user. For example, systems that perform multiple parallel searches on multiple databases and only mention the most promising few databases as sources for further investigation to the patron will become commonplace. Systems will optimize response time to the user and effectiveness of results rather than consumption of computing resources.

Tomorrow's systems will be more diverse than today's, presenting a greater range of different personalities adapted for different classes of users. They will provide help not only in using information sources, but in selecting them. Part of this future will come by way of a more integrated view of multiple databases, and part just by reorienting the user interface to recognize that the searching process also implies a selection of sources to search. It is interesting to note, in this connection, that recently developed systems designed for the networked information environment such as WAIS incorporate this selection of sources as an integral part of the standard user-system interaction.

These next-generation systems will support not only "active" searching, in which the user connects to the system to find information, but also a wide range of current awareness services, in which the user simply connects and is told what is new that might be of interest, or in which the system even contacts the user (or the user's workstation) through electronic mail or other mechanisms when something new and particularly interesting to that user enters one of the system's databases. Clearly, such systems must support non-anonymous access and make extensive use of long-time horizon user-system history to customize interactions through information gathered about users' preferences and past history.

We will not simply modernize the catalog in the new generation of systems; we will conceptualize and create a new class of information systems that include, as a part of their function, the traditional functions offered by card catalogs and automated card catalogs. One challenge will be coming up with an appropriate name for this new generation of information access and retrieval systems that reflects the extent to which they go beyond the traditional automated catalog both in scope of contents and in functionality. It is time to stop calling them online catalogs—the term itself is unduly limiting as we come to reconceptualize them.

ACKNOWLEDGMENTS

I would like to thank Michael Buckland and Michael Berger for many discussions about online catalogs and their future; Richard West for discussions of the relationship of library automation to university infrastructure; Cecilia M. Preston for comments on various drafts; and Nancy Gusack for her editorial assistance.

REFERENCES AND NOTES

1. Indeed, as the CLR study, *Users Look at Online Catalogs*, showed, the great majority of library users who had not yet encountered online catalogs were also enthusiastic about the prospects the new technology offered.
2. In fact, in the United Kingdom many card catalogs did not offer subject access, and I vividly recall a visit by a group of British librarians in the early 1980s, who seemed astounded that we were wasting resources on online subject access.
3. Ray R. Larson, *Users Look at Online Catalogs: Results of a National Survey of Users and Non-Users of Online Public Access Catalogs* (Berkeley, Calif.: Division of Library Automation and Library Research and Analysis Group, Office of the Assistant Vice-President, Library Plans and Policies, Univ. of California Systemwide Administration, 1982).
4. Ray R. Larson. *Users Look at Online Catalogs: Part 2, Interacting with Online Catalogs: Final Report to the Council on Library Resources* (Berkeley, Calif.: Division of Library Automation and Library Research and Analysis Group, Office of the Assistant Vice-President, Library Plans and Policies, Univ. of California Systemwide Administration, 1983).
5. The University of California, Office of the Executive Director for Universitywide Library Planning, *The University Of California Libraries: A Plan for Development 1978-1988* (Berkeley, Calif.: Library Plans and Policies, The Univ. of California, 1977).
6. Edwin B. Brownrigg and Clifford A. Lynch, "Beyond the Integrated Library System Concept: Bibliographic Networking at the University of California," *Proceedings, Second National Conference on Integrated Online Library Systems, Atlanta, Georgia, September 13–14, 1984* (Canfield, Ohio: Genaway and Associates, 1984), p.243–52.
7. Clifford A. Lynch, "The System Perspective," in *The Evolution of Library Automation: Management Issues and Future Perspectives*, ed. Gary M. Pitkin (Westport, Conn.: Meckler, 1991), p.39–57.
8. Current awareness services were familiar to users of systems like Dialogue in the 1970s, but these people were typically professional search intermediaries. In 1980, most patrons did not use a computer often enough for current awareness to be successful as an end-user service.
9. Clifford A. Lynch, "The Use of Heuristics in User Interface for Online Information Retrieval Systems," speech delivered at the ASIS annual meeting, 1987.
10. This applies to those database choices accessible through the MELVYL user interface, as opposed to through a remote terminal session gatewayed through the MELVYL system. When accessing a remote system, the user must use the native user interface of the remote resource, which can still be a problem. There are a number of unfriendly systems on the network, and some have a substantial learning curve. Some are radically different than the MELVYL system and confusing to MELVYL users because the user's existing experience with the MELVYL system is a treacherous guide and does not extend well. Also, network terminal emulation protocols add further complications. New technologies such as the Z39.50 computer-to-computer information retrieval protocol promise to permit access to ever larger numbers of remote information resources through a user's familiar local interface. This will further emphasize the problem of resource selection rather than access mechanics.
11. Clifford A. Lynch and Cecilia M. Preston, "Describing and Classifying Networked Information Resources," *Electronic Networking: Research, Applications, and Policy* 2, no.1:10–22 (Spring 1992).
12. Clifford A. Lynch, "Networked Information: A Revolution in Progress," in *Virtual Libraries: Implications for the Research Library*, ed. Brett Sutton and Charles H. Davis (Urbana-Champaign, Ill.: Graduate School of Library and Information Science, University of Illinois, 1992), p.12–39. ■■