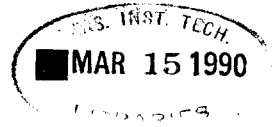# COMMUNICATIONS
FORUM

THE ELECTRONIC LIBRARY:

VISION AND IMPLEMENTATION

October 26, 1989

Seminar Notes

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

COMMUNICATIONS FORUM

THE ELECTRONIC LIBRARY:

VISION AND IMPLEMENTATION

October 26, 1989

Seminar Notes

Vinton Cerf, Corporation for National Research Initiatives

Patricia Battin, Commission on Preservation and Access

Steven Lerman, Professor of Civil Engineering, M.I.T.

Mark Kibbey, Project Mercury, Carnegie Mellon University

Gregory A. Jackson, Harvard Graduate School of Education

John Garrett, Copyright Clearance Center, Inc.

Brian Kahin, Kennedy School of Government, Harvard University

Antonio J. Botelho, M.I.T., Rapporteur

This session of MIT's Communication Forum dealt with electronic libraries. Brian Kahin, the session organizer, remarked that the topic is timely because there are several converging factors. We are at the point which visions of the electronic library are confronted with implementation problems. One of the most important developments is the success of library automation, particularly the bibliographic utilities. Another is the microcomputer revolution in the end-user community, which has allowed the user to learn search strategies and make use of sophisticated interfaces. The third factor is the evolution of networks, local and wide area networks (Bitnet, Internet); including the vision of a national research and educational network. What is unique about these networks is that unlike commercial networks, pricing is based on availability rather than use. A fourth factor is the emergence of optical storage, especially CD-ROM, a mass market technology and publication medium which is well suited to libraries. A factor of a different nature, is the ferment in federal information policy: the recent OTA study Informing the Nation, legislation to reauthorise the Paperwork Reduction Act, and profound policy shift within OMB that seems to favor greater government involvement in electronic publishing.

Kahin announced that the panel would seek to address both the opportunities and problems facing the electronic network. Opportunities ranging from new insights in distributed intelligent systems to new policies for preserving information and to new opportunities in the development of scholarly communication. Problems are to be found in the familiar areas of financing, institutional inertia and structure, and intellectual property.

The first speaker, Vinton Cerf, VP of the Corporation for National Research Initiatives (NRI), Washington, D.C.; worked on the original Arpanet, taught at Stanford, and joined DARPA in 1976. In 1982, Cerf joined MCI where he developed MCI Mail and joined NRI in 1986.

Cerf began by explaining what NRI is about. NRI is a non profit organization formed in 1986 which believes that information technology, networking and computer science can come together now to be fashioned into an information infrastructure. The US is ready for and desperately needs an information infrastructure to harness its computer power and communications infrastructure. There are three reasons for that: productivity, competitiveness and wetware enhancer. This last point is particularly important, because computers are the only things that can emulate the functionality of thought. Other infrastructures already in place are national and interstate highway systems, electrical power generation and distribution and telecommunications. All of those are economic engines enabling a substantial amount of business to happen. NRI is engaged in a number of related different projects: high-speed networks (NSF/DARPA), digital library system, knowledge bank, and electronic transaction framework.

Cerf then focused on the Digital Library System. First, he turned to information. Referring to Exhibit 1, Cerf pointed that one kind of information one might want to include in a library is printable documentation, which has fixed content and presentation format. The next possibility would be information that has fixed content but which format is not necessarily fixed, you might need to adapt the format depending on the medium you are

using. As one moves to the left of the diagram in Exhibit 1, you increasingly need computer-assistance to make the material usable.

In the middle part of Exhibit 1 there are databases, which requires computer support to query. Even more computer assistance is needed to make use of sensor data. As you move out toward the software end of the spectrum you have executable programs, which requires a machine to interpret them, and finally knowledge structures. These are ways of capturing, for example the kind of textbook knowledge we have now, in a machine-manageable knowledge structure which the machine can use to reason with. An example of that can be a computer-assisted design workstation used to design Ga-As integrated circuits. The concept of a digital library covers a very broad spectrum of information.

At the bottom level, Cerf continued, the most important concept is that the DLS idea depends on the possibility of inserting workstation/ computer power between the user and the information. The chosen paradigm for operating on information in the DLS is called knowledge robots or knowbots™M. Rather than using a conventional query this paradigm engages user interaction in order to extract from the user the types of information in which he/she is interested and to fabricate from that exchange a working program, the knowbot, and launch the knowbot into the network (See Exhibit 2). The knowbot has the ability to persist (over time, over machines, over reboots), differently from the query, and the capability to clone itself since the information will be spread in several databases, a multiplicity of information sources.

What would a DL system look like? (refer to Exhibit 3) First, there has to be a way of getting things in and out of the DLS, and where it enters it could be quite remote from the places where it might be manipulated, stored or viewed. One has to have a communications system, a network like Internet. Once the information has been brought into the system we have to figure out: what is it? does anybody own it? are there royalties associated with it? are there access controls to restrict use? are there manipulation restrictions? how can it be found again? That means there should be one or more registration servers in the architecture. Then, where should we store this piece of information? Today, this function is performed by people responsible for indexing and cataloguing.

Once this is done, the information will have to go into "knowbotic"-proficient database servers, since this information will be retrieved by knowbots. In addition, the DLS will have to be compatible with existing databases. Further, the system will have to incorporate the ability to account for use, for such a system cannot exist in the long run if it is not economically feasible. Next, the system should be able to incorporate Personal Library Systems (PLS), which someday will be a part of a user's home library. The idea is that the initial query would be done at the PLS, and if information was not found locally, the query would be extended to outside libraries in the architecture.

Looking further inside the DLS system (Exhibit 4), Cerf pointed out that the system will need a Knowbot Operating Environment. Making an analogy with hotels, he remarked that what the system describes is very much like a hotel service. Next, referring to Exhibit 5, Cerf pointed out that it represented a slicing down into the network, with the vertical lines

representing boundaries between computer systems. What we end up with is a collection of knowbots in different computers sitting in a common operating environment (KNOE). More critically, Cerf continued, we should not make the mistake of overlooking ways for interchanging queries and contents with other existing DLSs. Will be possible for knowbots to move back and forth across the systems?

Cerf then noted that as the discussion of the issues arising in the construction of a DLS evolved, NRI concluded that the information in the system needed to be protected in some way or another. Referring to Exhibit 6, Cerf, said that we have to imagine every piece of information as surrounded by a layer of software. You can not get into the object itself without passing through the software that will interpret messages that are asking to do things with that data. Potentially, the software can actually act as a guard. In addition, the import data function says that every object comes with its content and some information about where it came from, whether or not anybody owns the content and if so what sorts of royalty arrangements and access restrictions are there. That information is available to software which will interpret queries. Referring to Exhibit 7, Cerf commented that in the DLS the user starts with information held in a private space and then you could move that information into other parts of the library, represented by the annulae in Exhibit 7, which have increasing degrees of accessibility to the broader community. Very much like publishing, you can decide to which community will you make the information available.

Turning to the issue of intellectual property, Cerf stated that one might informally model objects in the DLS (Exhibit 8) as having "buttons" that cause actions to happen. The object interpreter determines which buttons have been "pushed" and by whom to make a decision about whether or not to honor the request.

Next, Cerf expanded on what a user interface might be like in the future DLS. He sees a 3D view of the library with intensive use of icons. The objective is to create an electronic analogue to the paradigm used to work with physical material. It is also inescapable that to make such a system acceptable one will have to use natural language interfaces.

Cerf briefly mentioned the future importance of compound documents: multifont, graphics & imagery, sound, animation, standards, editors, and hypermedia concepts. Interneting will also be important, because there will be more than one network, and remarked the growing importance of fast packet switching, running in the gigabit range. FPS means for the DLS that you can carry substantial amounts of information (bitmaps, video) in a short time. Furthermore, DLS should also incorporate active mailboxes, linked up to the knowbots, so that queries could be done without you having to be there. Another component of DLS, Cerf referred to is retrieval technology. Authentication technology, Cerf remarked, will be critical for DLS. If someone puts information into a DSL, let's say a scholar with the expectation of making his/her work to colleagues it is important that this information be not modified and that its integrity protected and maintained.

Cerf concluded his presentation with an overview of the current status of the DLS project. The final architecture document will be published this year, work has started on the interfaces and specifications, and analysis of copyright issues is underway. The next steps will be to design and specify the systems and identify potential standards.

The next speaker, Pat Battin; B.A.Swarthmore, M.S.L.S. Syracuse; pioneer of integration of computer and libraries at Columbia University; has since 1987 been the president of the Commission on Preservation and Access, Washington, D.C.

Battin's presentation was from a library perspective, skeptical of the previous visions. She said that for years, the visionaries have talked about "Libraries Without Walls;" she thinks perhaps those referred to as visionaries have been guilty of indulging in Vision without Walls: speculating on the basis of technological capability rather than intellectual significance and fiscal reality. Our society created the library to protect the important right to access to information. Will that right survive the Electronic Library? Or will the costs of technology be so great that we will settle for short term profits at the expense of intellectual productivity and quality. For example, let's take the year 2010 -- the end of our twenty year effort to reformatting three million rotting volumes because of acid paper. In 2010, scholarship, libraries and publishing will be completely transformed, with everyone -- students, scholars, and librarians (who will still be around) -- sitting in front of powerful workstations, linked around the world, sending and receiving both textual and non-textual data, images of a quality higher than the original, voice, music and other forms of multi-media not yet known to us. Vast amounts of text and image will be stored optically with rapid, inexpensive, bound volumes available on demand. Extremely sophisticated artificial intelligence software and expert systems will guide the user to the desired information, keep track of the cost-accounting for the use of such information, and provide custom-made packages of selected information to be routinely digested by the individual consumer.

Next, Battin asked what's wrong with this picture? According to her nothing except the fact that although this is an eminently plausible and admittedly conservative scenario from the technological perspective, most of the nations's private colleges and universities and well over half the public institutions filed for Chapter 11 in the mid-nineties in a futile effort to pay for the wave of costs generated by the enthusiastic and indiscriminate adoption of information technology in the exuberant eighties.

Battin then said that our visions of the capacity of technology to transform the processes of information generation, storage, dissemination, and use are realistically bounded by the human capacity to organize and manage complex systems for those functions, to create realistic and workable financial formulas for funding and controlling costs of widely decentralized use of information technologies, and to tailor the format of information storage and delivery to the extraordinary range of intellectual activities. So far, we have refused to admit in a spectacular display of obtuseness that our 19th century organizational structures and financial formulas will not work in the 21st century.

Battin remarked that it may well be that the weakest link in this technological vision of the future is the human brain, based on concerns: 1) the conceptualization, organization, and management of the standards, networks, and other functions necessary for convenient access to information; 2) the capacity of the human brain, as an integrative USER, of the enormous and continuing onslaught of information pouring from these new pipelines.

She cited a recent article in the Wall Street Journal, which remarked: "First they get us to

4

buy computers so we can get more information. Then the computers give us more information than we can ever read. Now they plan to sell us products that sift through all the information to give us what we really want to know." The significant point is the subtext of "sell" and "sale" running through this quote.

In a recent discussion, a colleague groused about the conventional dubbing of every technological change as a "revolution" as in the "computer revolution." His ill humor, she described, sent her to the dictionary where she found the following: "Revolution: sudden, radical, or complete change; basic reorientation and reorganization. Transformation: to change completely or essentially in composition or structure." The computer is not the revolution -- the truly technological revolution for librarians -- and other information specialists -- is the new capacity to generate, store, disseminate, and use information in entirely different formats. The format can now be tailored to the function. Format, in Battin's vision of 2010, becomes a tool rather than an icon.

Battin stated that this concept was critical to her vision of the future library: there will continue to be a broad spectrum of distinctive uses of information which will require differing formats, often by the same individual. The new ways will not supplant the old. The process of transforming the 19th century concept of the library based on print formats, stable research domains, and an orderly publishing process into the 21st century concept of the electronic library utilizing new technological capacities to enhance and extend the traditional functions will require revolutionary concepts for funding equitable access, for preserving and maintaining collective intellectual output, and for providing the appropriate balance between intellectual property rights and broad access to information in the public interest.
On the transformation and impatience of visionaries and technology specialists with the apparent slowness of large comprehensive libraries to move rapidly from 19th century infrastructure into the abyss, Battin commented that it is much easier to create than to transform; it is much easier to introduce new and ever-expanding technical capacities than to transform old ways into new habits, a process invariably accompanied by disruptive diversions in one's productivity and reluctant relinquishment of cherished traditions. The powerful mythology of the academic culture built around the book as icon presents an obstacle to overcome.

Battin's next statement was that the critical vision to have is not a scenario of the wonders of technology, but strategies for accomplishing a truly integrated "knowledge management" system in which supply and demand are evenly matched, a system which accommodates many formats and many functions. In a truly decentralized system within a centralized infrastructure, there is no economy of scale, a circumstance which truly challenges the information providers. We have our challenge -- how to manage technology in our own best interest.

The technology mavens are guilty of the same historic sin of librarians and to some extent, scientists and scholars, in their assumption that "one size fits all." And that we can confidently proceed, on the basis of assertions, rather than documented assumptions, that what is good for one information-seeking individual, is good for all; that what is good for one discipline is a model for all other disciplines. It is either intellectually or fiscally

5

responsible to blanket the universe with indiscriminate technical power, regardless of need, especially when we have to shift limited resources from one activity to fund another. Example: The bulk of the world's literature continues to be published in print format. Does it make sense to continue to divert scarce institutional resources to high-speed networks and expensive software when our best research libraries are now capable of collecting only 5 percent of the published literature; a decade ago it was 12-13 percent.

Battin pointed out that, despite the fact that scientists are the most heavily wired, the cost of scientific serial publications has heavily skewed the expenditure of acquisitions funds and seriously reduced the capacity of libraries to purchase materials in other disciplines, CD-ROM products, and remote access to external databases.

Space and maintenance costs for housing large, infrequently used book collections are growing exponentially. Yet we continue to shelve indiscriminately on prime real estate all materials regardless of use rather than seriously investigate alternative means of knowledge storage, whether it be remote warehouses (still expensive), shared optical storage with other institutions, or microfilm.

At the present time, Battin continued, we are bogged down in a destructive cycle of journal publishing. Knowledge is generated in the university, given to commercial publishing firms, peer reviewed, often for little or no recompense, and sold back to the universities at increasingly exorbitant costs. A number of plausible scenarios for regaining control of this process by mounting a cooperatively supported electronic journal distribution system have been articulated, but nothing happens. Why not? Why do scientists, who admit that the "real communication" goes on via e-mail and conferencing, continue to insist that the printed journal is critical to their research?

In the analysis of Harold Shapiro, President of Princeton University, colleges and universities have been able to keep up with the inflationary aspects of the labor-intensive activities of research and teaching, but traditional funding formulas are not adequate for financing the complexity spawned by technology. It is true that hardware costs continue to decrease, and that more and more power can sit on the desktop; however, all that power may not be needed by every individual and software costs will continue to increase with the power of hardware and the capacity of high-speed networks. Despite the many advantages of technology in reducing the labor-intensive costs of information transfer, -- and unfortunately, those costs savings, at least from library automation, are behind us -- the generation, organization, editing and publication of knowledge remains a costly labor-intensive activity of the human mind regardless of the format.
The basic hard truth, Battin asserted, is that the benefits of information technology accrue to the user and have never been in the university's budget. Where is the money to come from? And what are we doing to our traditional educational values and the concept of equity of access to information if we continue to rob Peter to pay Paul.

Battin went on to state that her vision of 2010: if we are to achieve the promise of technology we must embrace both intellectual and financial reality, preceded by a painful re-evaluation of our most traditional assumptions. The electronic library can only come about if we view the scholarly information function from a system-wide perspective as

seamless web no longer separable into autonomous divisions of the library, the computer center, the telecommunications facility, research and instructional support services, and the publishing community. The reality is that our traditional approaches, based on assumptions developed in a pre-technological era and not amenable to these new interdependencies, pit each of these groups against each other in an unproductive competition for institutional and external funds. The resulting fragmentation of financial resources is a continuing, pervasive diminution of access, services, and technical capacities across the system.

What do we have to do to make real progress in transforming our libraries in an intellectually and fiscally responsible fashion? Battin advanced that the first realization is that librarians can't do it alone. The changes are so fundamental that they require equally fundamental changes in the traditionally autonomous relationship among universities, compartmentalized organizations within universities, and budget formulas and procedures inherited from the nineteenth century. There are four areas that need to be transformed.

First, information technology. Battin argued that the cost-effective use of information technology requires a budget process which provides for three major categories of expenditure, each one of which carries with it an exigency unknown in an environment in which professorial sabbaticals are funded by simply eliminating the course offering from the affected semesters, filing and cataloguing backlogs can be tolerated to reduce staff expenditures, and institutional budgets can be balanced by committing meager discretionary funds on a one-time basis to a series of competing capital needs. Dependence upon information technology implies a commitment to a budget process which supports: 1) capital investment on a regular, recurring cycle; 2) recognition of increasing operating costs during the lifetime between upgrades; and 3) a series of incremental costs as new technologies generate new complexities, hardware and software requirements, and enhanced staff talents.

Second, the scholar at the workstation. In this regard, Battin argued that the scholarly use of technology has occasioned a fundamental change in the relationship of the information provider and the user. Traditionally, information requirements and habits were essentially shaped by the constraints of centralized systems of libraries and computer centers. In many instances, the course of research and instruction has been largely dependent upon the availability of information resources and services determined by the central allocation of institutional funds and decisions regarding the use of those funds by librarians and computer specialists. The increasing decentralization of technology and access to information to the scholar at the workstation now encourages the customization of inquiry, search and response, the inexorable process of scholarly specialization, and the seemingly infinite capacity to generate hitherto unknown costs.

Third, Battin, proceeded to provide a clear-eyed recognition of the costs of information services in the electronic library. An increasing demand for new and diversified services has accompanied the decentralization of technology. Despite the sophistication of the user, each new generation of hardware and software generally brings with it a quantum leap of complexity to be mastered. In the same manner, new storage and dissemination formats as well as the growing number of databases and knowledge resources require individualized assistance for effective use. The development of expert systems to guide individuals to appropriate information systems is a slow and costly process, lagging far behind the

galloping capabilities of computing and networking hardware. Cost is a real issue here: Instructional software for one course can run as high as several hundred thousand dollars.

And fourth, she concluded, there are significant changes to be made in the publishing community. The use of technology by the publishing industry has radically changed the traditional relationship between the scholar, publisher, and institutional customer. Unfortunately, economic advantage has not been one of the changes. The control of American scholarship by foreign conglomerates has also introduced a menacing change in the costs of access to scholarly publishing. Although the technology has existed for some time to provide electronic transmission of knowledge, the actual implementation has been slow to materialize. Why? Probably because the economics of electronic publication of scholarly materials are largely unknown. The paradoxical properties of information technology provide both the capacity to control access to information as well as the opportunity for indiscriminate copying. Many publishers of both print and software formats are urging the use of contract law, rather than copyright law, to control access to information in the belief that the fair use provisions of current copyright law do not provide adequate profit margins. In essence, the publishing industry appears to be in chaos, as it seeks to maximize profits on its paper formats while attempting to develop new markets for electronic products. Since the seventies, the rule, for both print and non-print products, is a complete disjuncture between cost and price.

In the electronic library of 2010, Battin stated, we will need to effect a series of changes. First, if scholarship is to flourish we will have to accommodate different formats for different functions, in the same discipline and in the same person: for example, e-mail, telephone, fax, overnight delivery, and regular mail. Second we will have to know much more than we do now about individual information habits, i.e. who needs electronic publishing and who needs hard copy; the relationship between scientists and printed journals; and the pattern of use of indexes and abstracts, such as the Science Citation Index. Third, we will have to develop new concepts to protect the principle of equity of access in an environment that considers information as a product rather than a public good. Fourth, we will need to manage distribution systems which respond to the lowest common access denominator so that information can be transmitted according to the level of reception available at the scholar's location. Fifth, we will have to agree on a workable balance between intellectual property rights and affordable convenient access to information in the public interest. Finally, we will have to build new interinstitutional fiscal and management structures to share the archival obligation in the national interest -- to build a critical national collection to provide a context for local physical collections and access alternatives. In this regard, Battin concluded, we have to look with healthy skepticism on partnerships with the federal government and the track record of reliability.

Battin concluded by saying that the electronic library of 2010 will have walls and books and people -- it will also have vast dissemination and retrieval capacities -- both robotic and human -- to the universe of knowledge in many formats and configurations. It will be the central infrastructure of the university, organized in a manner based on generic function, bearing no particular relationship to the unproductive, nineteenth century pattern of organization currently existing, a pattern which separates information functions by historic tradition and format rather than the seamless web required to serve the 21st century user.

8

The staff will include peerless managers, recognized in the university as equivalent in responsibility and talent to the senior professoriat, possessing vision, creativity, imagination in abundance, and adept in dealing with ambiguity and managing the "connectedness" of a decentralized technological environment. They will be assisted by disciplinary specialists with knowledge management and technical skills, and a host of technicians to perform bibliographic, clerical, programming and other service functions required by a comprehensive information technology operation.

Battin concluded that above all, since in a technological environment, the one constant is change, the electronic library of 2010, will have the capacity to evaluate and respond to change in an informed manner which promotes and encourages the fundamental missions of research and instruction in our society.

The next speaker was Professor Steven Lerman, Civil Engineering Department, MIT, former director of Project Athena and director of the Intelligent Engineering Systems Laboratory. Lerman's vision of the electronic library seeks to tie together all the knowledge out there. He pointed out to disparity between what is conceptually possible and what is actually happening in the area of information technology (IT). The reasons for this mismatch are cost and the fact that technology often does not fit the task.
Today's knowledge system can be characterized by several factors: 1- a paper-based society; 2- text still is the dominant information medium; 3- the way we move information is by moving pages of printed text; 4- there is still a high cost and volume threshold for publications; 5- until quite recently the time needed for publication has been acceptable, and 6- in the current system there has been a small number of participants in key fields. Lerman vision of the 21st century already sees early symptoms of collapse of a paper-based society. First, non-paper based information, such as databases, are growing in volume. Second, existing legal and regulatory institutions are struggling in the area of intellectual property rights to deal with the emerging challenges, for example in software. Third, there is a decrease in reliance of research on published work. Fourth, there is an explosion of specialized and disparate electronic services, such as electronic reservation systems in the airline industry; banking and financial systems; and legal systems such as Lexix. Finally, there is a growing sense among scholars that they are overloaded with information.

Why the information revolution has not happened? Lerman thinks that the amount of stored information has increased spectacularly. The total amount of digitally stored information in disk drives and other rotating media equals the sum of everything that existed beforehand; every year the added increment is often bigger than everything that existed before, thanks to improvements in magnetic and optical storage. This technology area seems finally be coming of age. Information processing power with the explosion of desktop computers signals that if the technology is not ready yet, it will be soon. In data communications the technology has moved rapidly but the necessary infrastructure has lagged. ISDN networks have progressed very slowly in the United States. In electronic publishing, the technology which allows the sidestepping the formatting part of the publishing process is developing quickly. In information display, the quality of paper remains exceptional, and reading large documents on VDT displays is still difficult. Standards for information storage, retrieval and formatting have lagged advances in

technology and without standards investments in this area will not be forthcoming.

Why are libraries facing these financial problems? According to Lerman, first, libraries have couched their argument for resources of the national pie in the worst possible form. We have to use some approach nearly as compelling as that used for financing the highway system. Second, the transition to an electronic information system is not going to be cheap. If we want to become an information society we have to invest in infrastructure. The way to make these things happen is to make a convincing case that we need a pilot project. The short term strategy thus should be a series of large-scale pilot projects. A promising area is the working paper collection. If scholars are saying that they are increasingly relying on things circulated in conferences, on working papers, and less on published journal literature, we should target electronic information to these things that are being used. There should also be some retrospective conversion work that would help us understand this process better. Another area is electronic journals, which unfortunately have relied too much on the paper journal format, throwing away the comparative advantage of electronic media. For example, the concept of bundling papers together in an issue of a journal is an artifact of printing technology, and should be abandoned in the electronic journal. Papers should be released whenever they become available. Another promising area is hypermedia technology. However, the cost of cataloguing and cross-referencing documents for hypermedia is extremely large. Thus we will have to come up with some sort of automatic hypermedia technology to bring down these costs. A final area is a prototype of the library of the future.

Lerman then elaborated on the reasons why the gap between what is theoretically possible and what we are doing is so large. First, computer technology has been driven by computing needs of business and engineering, not by the needs of information infrastructure, causing the misfit with the librarian needs. There are also areas in which industry is not ready to deliver the technology yet. A second major factor is cumulative inertia. We have been using paper for a very long time. Third, standards are still lagging, although there seems to be a change in recent times. Fourth, experimentation is expensive and has relatively long payoffs. The fifth reason is organizational inertia. Organizations are deeply committed to current methods and have a vested stake in paper-based communications. It may be that the electronic library of the mid-term it will be one where a lot happens electronically but the end user converts it in paper. Finally, the existing funds to develop our knowledge infrastructure are inadequate to sustain even the current evolution of our knowledge system.

Mark Kibbey, Director of Library Automation at Carnegie-Mellon, was the following speaker. Kibbey began by stating the need to look for pathways to move into the future. Such projects may be the best way to assess both the financial and organizational issues and to help decision-making. A vision is necessary to provide guidance.

Kibbey went on to state that Carnegie-Mellon's Mercury pilot project looks at providing electronic full text information in computer science. Mercury is a project with a broad vision: independence of location, ease of browsing and retrieval, and vastness of coverage. Another parallel development of future library services at Carnegie-Mellon is the Library Information System 1 (LIS 1) to be followed by LIS 2. The LIS systems are somewhat

divergent from Mercury: one focuses on a broad spectrum of information for the whole campus and the other is able to provide a critical mass of full text information by limiting the discipline to computer science. Kibbey stated that since this talk would focus on the shorter term projects of LIS 2.

The idea is to create a pilot project. Past experience has shown that collaboration is crucial; an electronic library will be the work of many organizations. The business arrangements among the organizations must provide incentives for all. It has also been learned that faculty, librarians and publishers do not share a common worldview, do not understand each other goals insignificant numbers, and periodically design projects that do not suit each other's means. Thus, one of the things we need is better communication. Another lesson is that the economics of publishing are complex. There are a number of interested parties (authors, publishers, libraries, readers) operating in a stable environment and they are not interdependent. Payment for services is complex (subscription; use based); and control on distribution takes many forms (legal, technical, ethical).

A basic objective of LIS 2 is to expand electronic resources. One way to do this Kibbey continued, is to systematically enhance bibliographic records to improve access to information. The question here is what information should be added? A fair list might include: book reviews from CHOICE, plays in collections, papers in conference proceedings, individually authored chapters in books, illustrators and artists in exhibition catalogs, architectural illustrations, technical reports/abstracts, and journal articles. The idea is that records in libraries catalogs should be done at a more detailed level. The implications are that the infrastructure needs to be changed to allow individual additions be incorporated in existing systems, such as OCLC. At Carnegie Mellon, if everything were cataloged at the item level, the amount of material added in a single year would equal to the entire present book catalogue. To have a hope of doing this, campuswide planning, in conjunction with network and computer personnel, is essential.

Kibbey addressed the objective of experimenting with storing and displaying full text documents. The lesson learned was that in the traditional publishing model the book is the official representation of a body of knowledge. In the electronic publishing model it must be a computer representation (refer to Exhibits 9 and 10). A key step to be taken, before the electronic library becomes economically feasible, is to have the computer representation as the definitive one, with publication and paper as by-products. The next objective deals with the provision of a user interface which takes advantage of personal computers. The implication is that our current sophisticated high speed networks with PCs, workstations, etc. has all of them acting as dumb terminals. What is needed is a client-server relationship, with the interface software running on a microcomputer. This environment has to be integrated with the growing campus environment: network, multivendor, microcomputing. Furthermore, there is a need to recognize that there will not be a single system on campus as different departments will produce their own databases running on separate machines.

Given the campus environment, standards are critical. An electronic library will be the work of many independent organizations. The interfaces between the components must follow set standards. Mercury uses a combination of software protocols: DECWindows (X-

11

11), Z39.50, ISODE/TCP/IP, and NEWTON Search Engine. Any search engine supporting Z 39.50 should work in this model. Currently, libraries' search systems have different window interfaces and syntax, so in the user interest there is a need for a standard.

Kibbey went on to state that by the end of the millennium it will be cheaper to store books on facsimile images on hard disks than it will be to build new library buildings. The problem is that someone will have to match up the up-front conversion costs. The only solution is to do it in incremental stages. Also, balancing services is one of the toughest problems.

Where are we today? In prototype mode DEC or VAXstation can run across the internet. (Refer to Exhibits 11 and 12).

The next speaker was, Gregory Jackson, Associate Professor of Education at Harvard University, and one of the founding members of the Harvard Education Technology Center. Jackson has also worked extensively in the development and evaluation of policy in higher education and is currently special assistant to the Dean of Graduate Education at MIT.

Jackson said that the center of his talk would be turf in the electronic library. The point is that things in organizations are the way they are for reasons. The findings presented are based primarily on a project based at Harvard University, but which also include other institutions (Refer to Exhibit 13). The project sought to answer questions about how Harvard should go about handling the electronic library.

Jackson said that one of the empirical findings was that from the experience of several organizations one identifies five dimensions of practice (Refer to Exhibit 14). But all that variation is not exploited, and we can collapse three basic kinds of practices (Refer to Exhibit 15). What is striking about these practices is that the people's ideals are different from these (Exhibit 15).

Jackson went on to say that one can distill six main themes, consistent across organizations (Refer to Exhibit 16). In contrast, the discrepant themes are also present, the seeds of the more pervasive turf problems. First, relationships between libraries and data centers are sometimes collaborative, sometimes hostile. He said that his own study at Harvard happened to be a good illustration of this theme. Second, people think that support for electronic materials is sometimes a central obligation (University of California), and sometimes an individual. When it is not central, exclusionary policies often result. Third, electronic materials are sometimes managed as services, sometimes as collections. At Harvard, for example, paper materials and electronic-based materials are handled and indexed differently. Fourth, new formats for storing, retrieving, and especially indexing scholarly material such as "hypermedia" often grow independently.

Jackson then moved on to develop some of the turf questions involved in the image and implementation of the electronic library. There is a distinction between research and instruction, in terms of how do you think about what should be available, in what format, at what cost, with what kind of support. The point is that these things replicate problems that have not been solved with other kinds of materials. A second has to do with scholarly

and political boundaries. A third issue has to do with standardization or flexibility. For example, there are 4-5 different search engines that will work with Medline at Harvard. Next, Jackson tried to provide answers to three operational questions (Refer to Exhibit 17). The answer to the first question is central provision is to local provision as catalogues are to specialized findings. That means that anything looks like a catalogue that everybody uses (union catalogues and basic periodical indexes) is a central job and should be available throughout the system without cost to patrons; anything that looks like something that would be found in a few specialized libraries is something that is the local job to provide. The answer to the second question is that network offering is to local offering as general journals are to specialized journals. Thus, electronic materials commonly used by patrons at several libraries representing several disciplines (general journals) should be available through university-wide channels such as networks; specialized journals (e.g. Cell) should be provided locally. The analogy also holds for general and specialized databases. The answer to the third question is that restricted (includes charging people) is to public access as reserve is to stack material. If libraries place materials on reserve when local patrons' needs make this necessary, then libraries should restrict electronic materials when serving "outsiders" to avoid limiting "insiders" to less access than they would like.

Where does all this go? Jackson's answer as that the kind of vexatious problems in traditional materials still exist in analogies for electronic materials. So the argument is that this image of the electronic library as something that is "Central," we do not perceive where its boundaries are, has some elements of fantasy to it. Lots of old boundaries will continue and if the need arises we will invent other kinds of boundaries. Other distinctions are: between institutions and between clumps of institutions (society and individual institutions). There is a turf battle between social benefits and individual costs.

The following speaker was John Garrett, manager of market development, Copyright Clearance Center (CCC). The Copyright Clearance Center mediates between copyright holders and users, standing right in the middle of people who own information and those who want use it. The CCC has been developing mechanisms for coping with situations arising from these relationships that are addressing fundamental questions. Garrett remarked that innovations in this area come from corporations, not universities.
Garrett went on to say that one of CCC's main area of interest is in paradigms of use: people do not know yet what they are going to need in the future to cope with the enormous amounts of paper too fast too often. What we really need is a computer system to say: this is what I do, this is what I want you to do. And the only models are those based on print form. But CCC is convinced that what people will consult on a screen will be fundamentally different from what they consult in paper day to day. CCC is interested in understanding what people want to have when they have it on the screen. Who is going to use what? How often will people use what for what purpose; whether or not they read it on the screen; whether or not they print it out; what do they do with that printout; whether they download into their personal computer system; whether they create a private work product that integrates work from different sources. On the other side, representing right holders, CCC is very interested in knowing who is going to permit what uses for what purpose. Library community and publisher community ask the same questions. CCC is trying to bring right holders and users together little by little, getting authorizations one by one. At what cost is vital issue, publishers have no idea how to price these things. There

are some scary paradigms out there such as price per use. CCC proposal is simple. If a user wants to use a particular set of full text materials owned by a particular publisher; the publisher says that if you have a certain number of terminals that will access; it is going to cost X per year.

Garrett's observations CCC's role in standards included that the danger here is that with so many people trying to do something about it we will end up with 60 different essentially incompatible standards. CCC is trying to learn who is going to pay what for what. So far the economics are clouded by computer firms who particularly put a lot of money in educational projects. What would happen to universities if they had to pay for all the workstations they get? That is why the corporate model is very interesting to us, because corporations actually have to make money to pay for what they buy. For example, everybody knows that wordprocessors do not save much money and time. But there is so much out there that nobody wants to go back to the typewriter. Will the electronic library actually save anybody any money or increase productivity? Nobody knows. The economic modelling of this is key.

Garrett then expressed his fear of a government-operated program in this area. First, because of privacy. Second, big brother risks. The model CCC is looking into is one in which people pay for what they get. CCC has some interesting experiences. The best model is private practice law, completely dependent on computerized full text electronic information. The lessons are: costs are passed through and the more electronic forms you sell the more books you sell.

Finally, Garrett talked about some CCC projects. The long term one involves a major university molecular genetics research center. Here all major research information (200 journals and 25 books) will be provided electronically, as part of a four year project. The project has raised some interesting issues. The screen displays are great but there is no printer that is good enough to give for example a detailed image of an x-ray. Also screen resolution for some uses is not enough. A short term project underway is a national corporate center specializing in manufacturing literature. The idea is to put selected articles from 250 journals into an electronic system so that it can be sent to member corporations, to about 130 terminals around the United States. That will be in place by February 1990, and they will be doing that authorized by rights holders. Another project in a midwestern state has put PCs in the homes of victims of AIDS in early stages of disease. CCC is helping them to authorize in putting articles on AIDs management and research into that electronic system.

According to Garrett the programs vary widely in terms of size and content, but all are aimed at trying to understand where these process are going to take us in terms of use. He added his concern about the extent to which we are dragging thoughtlessly print models of use, of placement, of organization, of pagination, of illustration into this alternative world. It seems that we are, perhaps, dragging the worst of two worlds. In summary, Garret said he fell in the romantic-skeptic mode and that he is exceedingly skeptical about our ability to use these methods wisely and to increase rather than decrease human freedom.

In the ensuing question and answer period, answering to a question about negotiation,

Garrett said that the CCC established with the Association of American Publishers, a committee representing some of the major scientific and technical publishers. In relation to electronic material, the price will be negotiated individually.

Cerf remarked in relation to standards, that NRI had an experience recently with Internet in which NRI tried to a take an electronic publishing series available in ASCII format and substituted for Postcript. However, given the wave of objections NRI was forced to offer the series in both formats, because people were using one format for data search; pointing to the conclusion that more than one format will be needed. Cerf then asked Garrett whether a distinction can be made of the costs of different formats, so that those that are more manipulable could command a higher price. Garrett's answer was that manipulation is a critical issue for authors as for publishers and that technical standards are the least likely to be imposed.

The following question asked for comments on what happens to intellectual curiosity when you have to pay per access? Cerf responded that there is a public cost to information in public libraries. In fact, one of the parameters of designing a successful electronic library is to find similar means to subsidize that cost. Battin added that all libraries are subsidized. Is it possible to look at the public utility concept? Given the nature of the intellectual property system, there seems to be a possibility for a balance between treating information as a commodity and treating information as a public good. Cerf replied that some of the problems faced by the printed format will be easier to solve in the electronic library, such as photocopying. In the electronic format, the scary issue is the easy that copying can be done. Battin replied that before, in the print era, we had a choice, but now a lot of information is controlled. The problem now is to define what basic information services are included in the tuition and what value-added, discretionary services are available for additional fees.

The next question dealt with possible publishing conflicts of not for profit public organizations like NRC and AAAS. Garrett said that librarians do not make that distinction when complaining about the rising cost of publications. Kibbey remarked that the societies are as concerned as the for profit publishers. Cerf added that he was surprised when he found out that a large part of the scientific community does not operate on the basis of working papers like in engineering. Perhaps, this could become an acceptable practice for other disciplines, although problems of quality of publication and tenure mechanism linger.

The next question was: what will happen to the gap between the "haves" and the "have not"? For Jackson, in the case of the developed countries, the trickle down effect is more likely to be positive than negative. Once these electronic materials become available, and major corporations and research universities are likely to bear the costs of conversion, then the price will come down. The more serious problem has to do with less developed countries, where the underlying infrastructure to make use of these materials is likely to lag significantly. Garrett added that for many of the technical centers in developing countries they are more technically advanced in terms of the use of electronic information because the United Nations made a major point of subsidizing those purposes simply because the other literature is not available to them. Battin added that it is quite possible we will see

15

the development of "information haves- and have-nots" in this country. Jackson added that the "have not" problem can become worse as the "have not" become aware of the immense size of the gap, creating all sort of political pressure in the "have not" institutions. Battin replied that we do not know how the electronic library will affect productivity because we do not really know the relationship of research productivity to access to literature in print format.
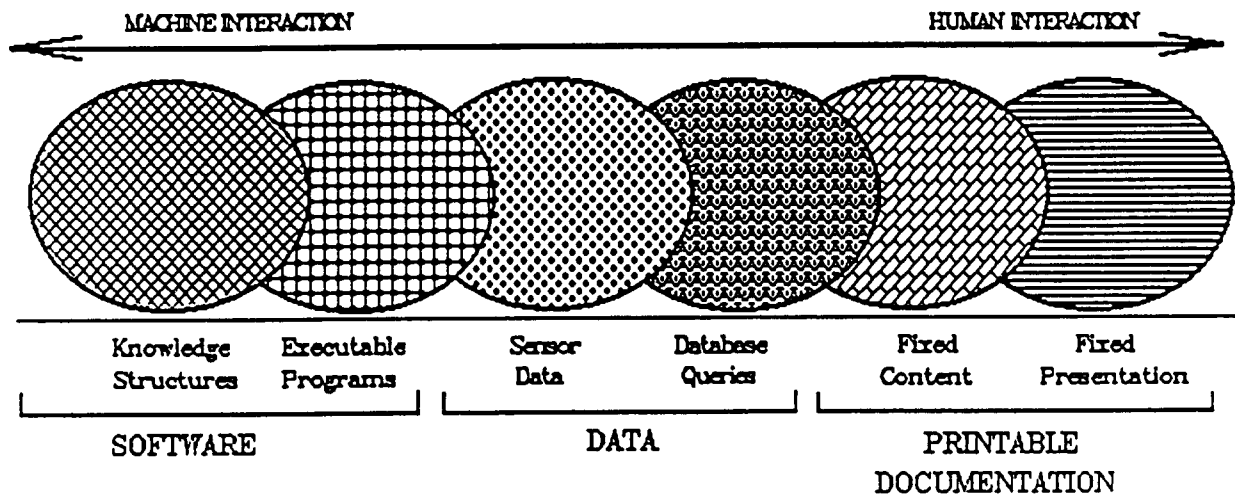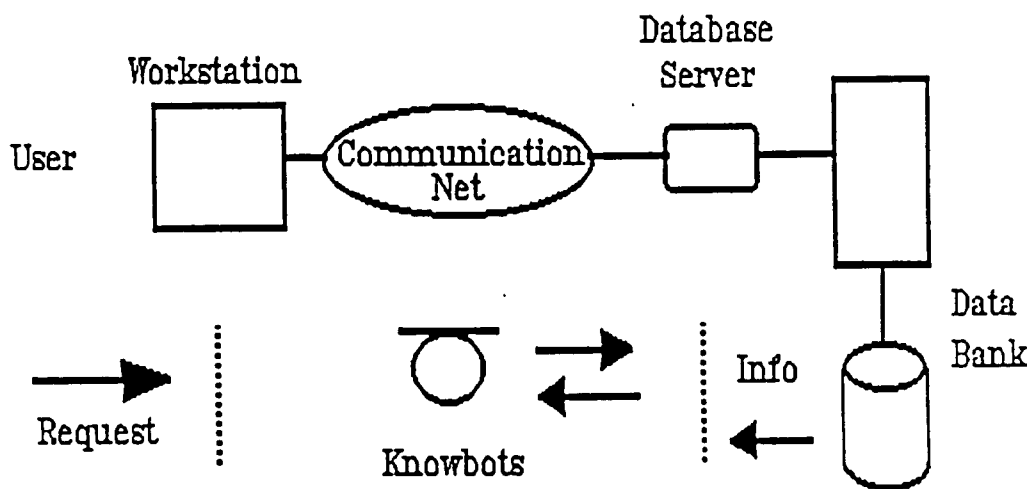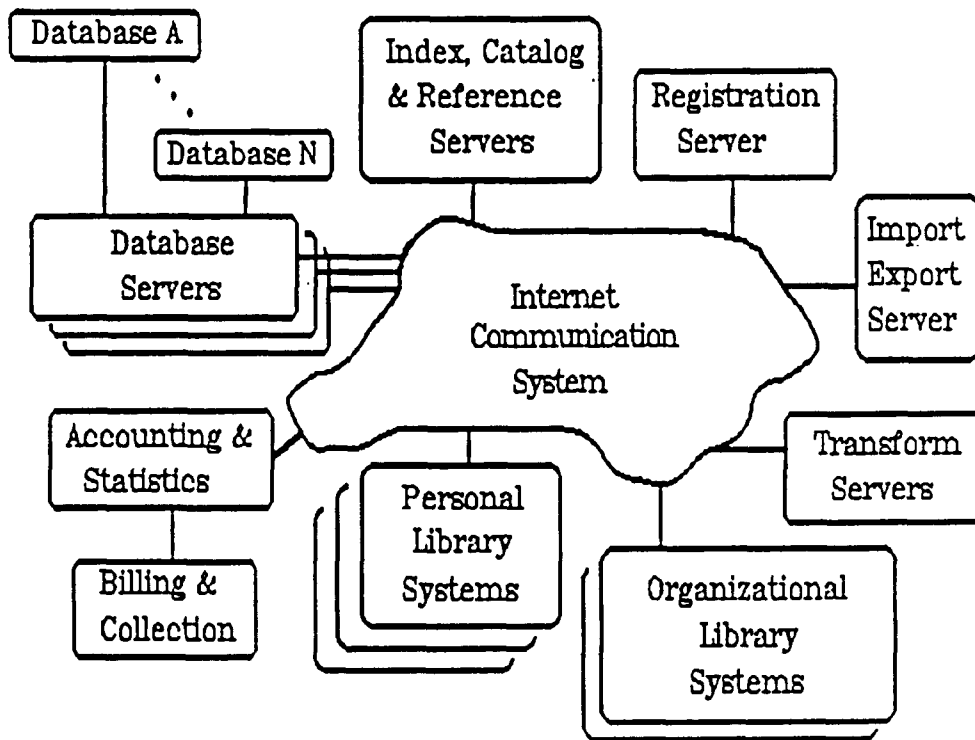
MACHINE INTERACTION   HUMAN INTERACTION

| Knowledge | Executable | Sensor | Database | Fixed | Fixed |
| Structures | Programs | Data | Queries | Content | Presentation |

SOFTWARE   DATA   PRINTABLE
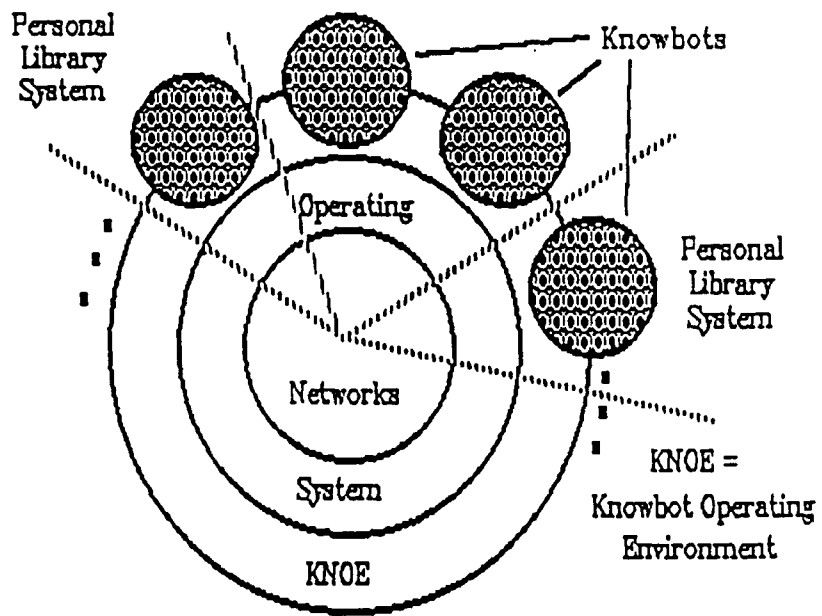   DOCUMENTATION

EXHIBIT I

Conceptual Organization

User · Workstation · Communication Net · Database Server · Data Bank

Request · Knowbots · Info

EXHIBIT 2

Database A

Database N

Index, Catalog
& Reference
Servers

Registration
Server

Database
Servers

Import
Export
Server

Internet
Communication
System

Accounting &
Statistics

Transform
Servers

Personal
Library
Systems

Billing &
Collection

Organizational
Library
Systems

EXHIBIT 3

USER AGENTS	LIBRARY CONTENT	OPERATIONAL SUPPORT
Doc. Edit. Doc. Srch. Import Export Icon Geography Acct. Transform	Doc. Reg. Obj. Srch. Library Statistics Index & Catalog Library Cartography	User Reg. Usage Reporting Diagnostics Backup

KNOWBOT OPERATING ENVIRONMENT	NATURAL LANGUAGE SUPPORT

PROGRAMMING LANG. SUPPORT	OBJECT REPOSITORY	MESSAGE TRANSPORT AGENT	TRANS- ACTION PROT.	TRANS- PORTATION PROTOCOL

PRESENTATION SERVICES	FILE SERVICES	TRANSPORT PROTOCOLS

OPERATING SYSTEM	DEVICE AND NETWORK DRIVERS

EXHIBIT 4

Personal
Library
System

Knowbots

Operating

Networks

System

KNOE

Personal
Library
System

KNOE =
Knowbot Operating
Environment

EXHIBIT 5

Message
Interpreter

Import Data

Object
Content

Object Model

EXHIBIT 6

private
space

annulus

object

Public Space

Object Library

The Annulus Concept

EXHIBIT 7   Figure 2

Buttons

(Inputs)

Message
Interpreter →

Object

(Outputs)
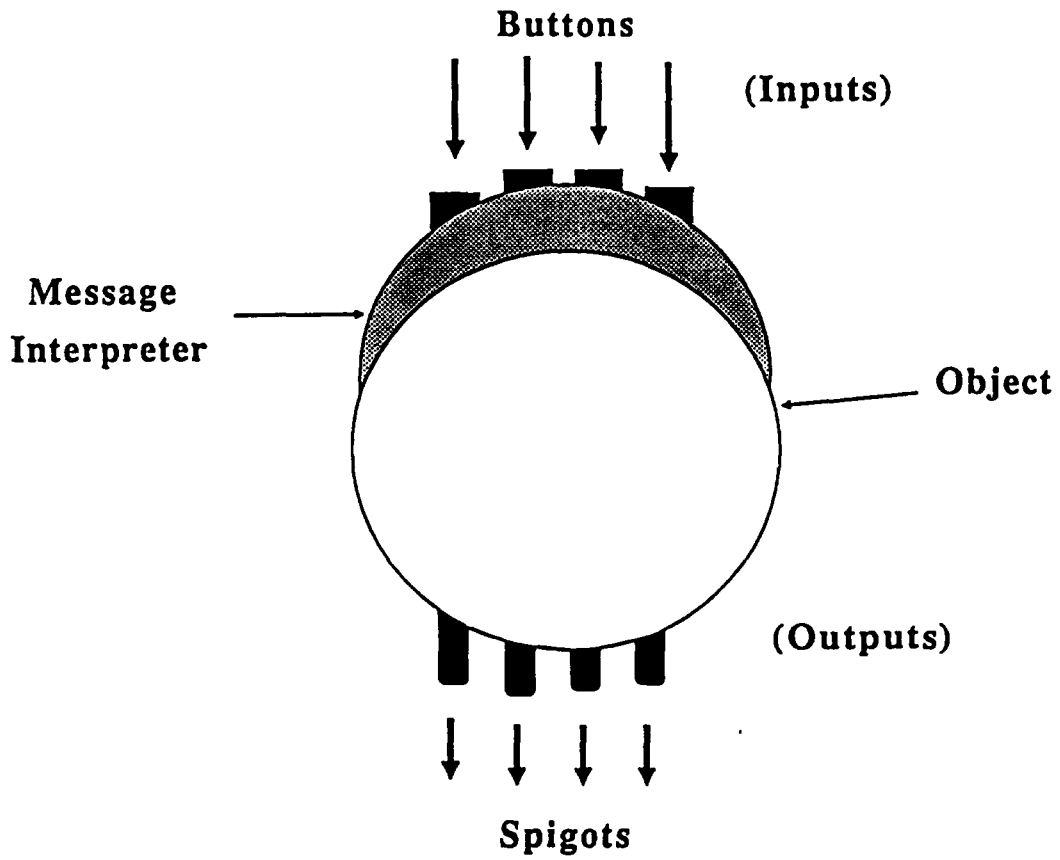
Spigots

Buttons and Spigots

Figure 4

EXHIBIT 8

Electronic publishing

EXHIBIT 9
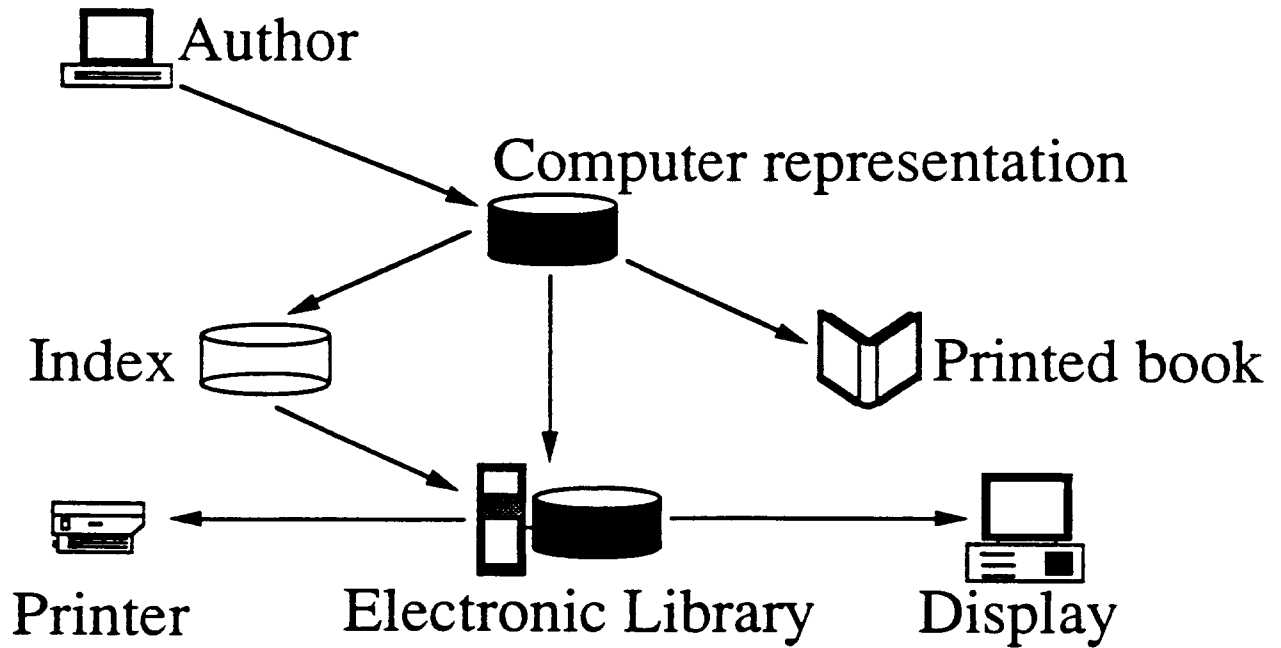
Traditional publishing and
the Electronic Library

Author

Book

Scanner

Index

Printer   Electronic Library   Display
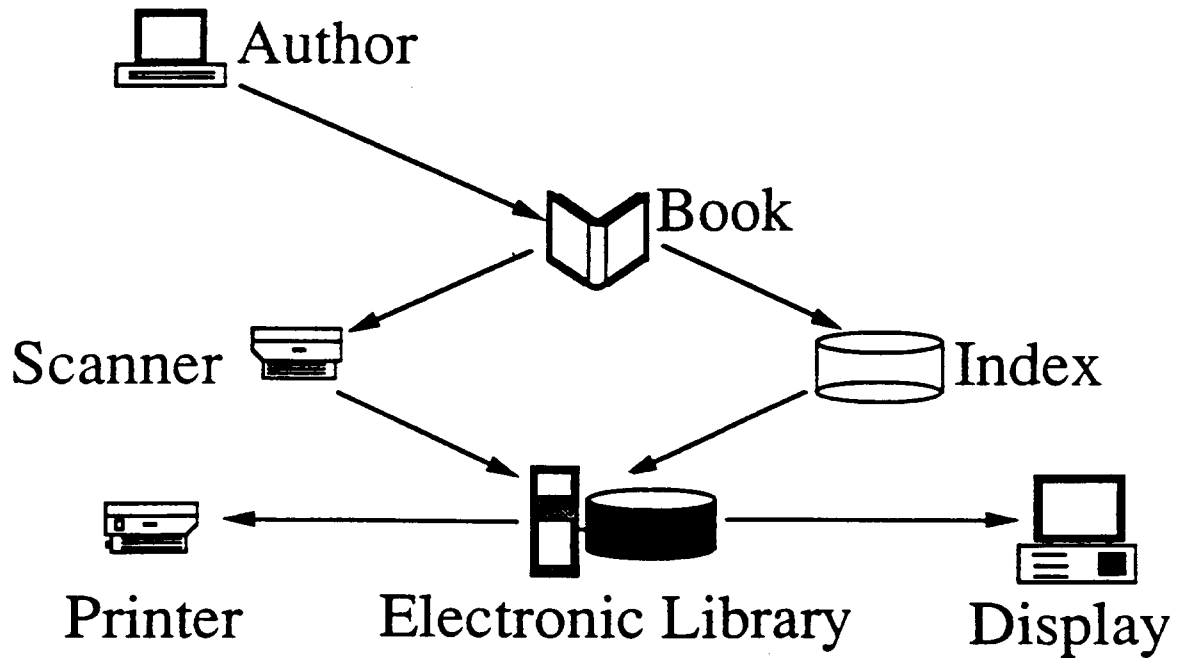
MERCURY

Building blocks

Carnegie Mellon

      Library information system
      Campus network & NSFnet
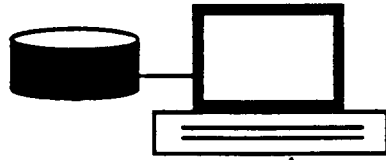   Andrew project
      Computer science
      Computational linguistics

OCLC

      Newton database system
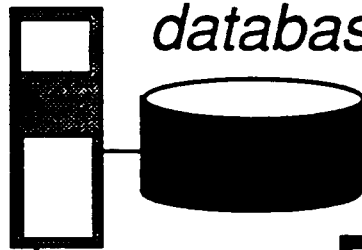   Z39.50 protocol development

Digital Equipment Corporation

      Workstation and server equipment
      X window manager
      Compound Document Architecture
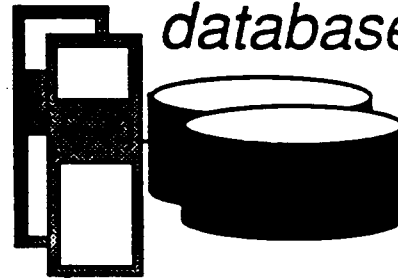
EXHIBIT II

The Mercury
Electronic Research Library
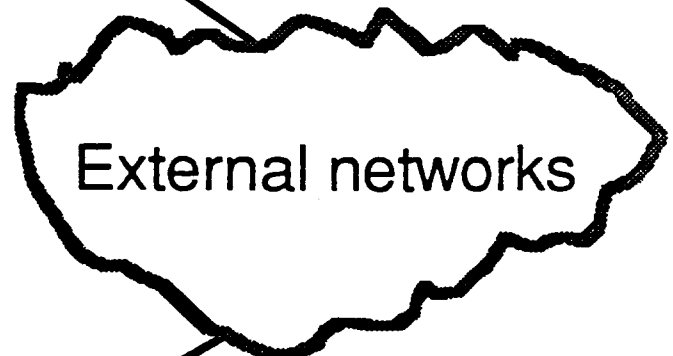
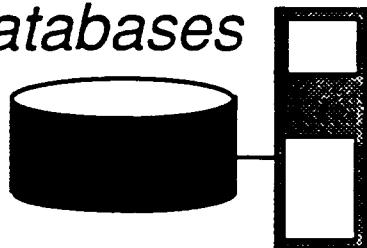*Workstation
with local databases*

*Departmental
databases*

*Library
databases*

*Reference
assistant*

External networks

*Remote
databases*

CASES

Harvard University

Education, Business, Law, Government, Arts and
Sciences, University Library, Office for Information
Technology

Stanford University

University of California

Dartmouth College

Carnegie-Mellon University

University of Southern California

Miami-Dade Community College

EXHIBIT I3

2 GJackson 10/26/89

DIMENSIONS OF PRACTICE

Retail . . . . . Wholesale

Is access to electronic materials acquired in bulk, or on an as-needed basis?

Local . . . . . Distributed

Is access to electronic materials at specific sites only, or anywhere throughout a library system or campus network?

Restricted . . . . . Free

Is access to electronic materials restricted, absolutely or through disincentives such as signups or charges, or is it relatively unconstrained?

Mediated . . . . . Direct

Do patrons themselves use electronic materials, or must reference librarians or other professionals be involved?

Index . . . . . General

Do electronic materials simply provide indexes and similar guides to traditional materials, or do they include full texts, raw data, tables, or other such documents?

TYPICAL CONFIGURATIONS

Most: Retail-Local-Restricted-Mediated-Index
(Librarian-assisted, patron-paid dialups)

Some: Retail-Local-Free-Direct-Index
(Selected CD-ROMs)

Few: Retail-Local-Free-Direct-Index
+
Retail-Local-Restricted-Direct-General
(CD-ROMs plus some network full texts at a fee)

Ideal: Wholesale-Distributed-Free-Direct-General
(Free network access to everything)

CONSISTENT THEMES

- Electronic materials are prevalent in university and college libraries.

- Certain arrangements have become conventional.

- Conventional arrangements follow an evolutionary path.

- Financing shifts gradually from variable to fixed fees, and from users to libraries to university core budgets.

- CD-ROM is a transitional technology.

- Research and instruction entail different arrangements.

OPERATIONAL QUESTIONS

- When should electronic materials be acquired centrally, as a system-wide responsibility, and when should they be left to local prerogatives and budgets?

- When should access to retail databases -- that is, those acquired over dialup or network connections at costs proportional to use -- be local, within a specific library, and when should it be distributed through a network?

- When should locally provided databases be available with minimal restrictions (as is quite possible with CD-ROM and similar technologies whose cost is not proportional to use), and when with price or other restrictions?

EXHIBIT 17

8 GJackson 10/26/89