DOCUMENT RESUME

ED 328 258 IR 053 415

TITLE	Beyond Bibliographic Data: Proceedings of the Library of Congress Network Advisory Committee Meeting (Washington, D.C., March 29-31, 1989). Network Planning Paper No. 19.
INSTITUTION	Library of Congress, Washington, DC. Network Development and MARC Standards Office.
REPORT NO	ISBN-0-8444-0657-0; ISSN-0160-9742
PUB DATE	89
NOTE	60p.; For Network Planning Paper No. 17--1988 conference, see ED 315 098.
AVAILABLE FROM	Cataloging Distribution Service, Library of Congress, Washington, DC 20541.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Bibliographic Databases; Computer Networks; *Full Text Databases; *Information Networks; *Information Retrieval; Information Storage; *Integrated Library Systems; Library Networks; National Programs; *Numeric Databases; Online Searching; *Research Tools

ABSTRACT

The seven papers in these proceedings focus on the characteristics and potentials of non-bibliographic databases in the context of library and commercial networks. Following an introduction by Sandra K. Paul, this report presents the individual papers: (1) "User Perspectives and Requirements: Creator of Non-Bibliographic Databases Has To Share with Others" (Rudolph M. Bell); (2) "Making Remotely Sensed Data More Accessible" (Charles E. Olson, Jr.); (3) "Networking and Access to Non-Bibliographic Databases: A Commercial Perspective" (Dennis McDonald); (4) "The Research Libraries Group (RLG) Assessment of Non-Bibliographic Information Needs" (Constance Gould); (5) "Summary of Three Non-Bibliographic Operating Entities" (C. James Schmidt); (6) "Beyond Bibliography or Creating the Rosetta Stone for the 21st Century" (Kenneth Dowlin); and (7) "Non-Bibliographic Databases in the Network Context: Meeting Notes and Discussion Comments" (Ronald F. Miller). A summary of the business session and a meeting agenda are included. (MAB)

NETWORK PLANNING PAPERS

No. 18 *Connecting the Networks.* Proceedings of the joint Library of Congress Network Advisory Committee and EDUCOM Meeting, December 5–7, 1988. (Washington: Library of Congress, 1989). 102 p. $10.00.

No. 17 *Intellectual Property Issues in the Library Network Context:* Proceedings of the Library of Congress Network Advisory Committee Meeting, March 23–25, 1988. (Washington: Library of Congress, 1988). 98 p. $7.50.

No. 16 *Intellectual Property Rights in an Electronic Age:* Proceedings of the Library of Congress Network Advisory Committee Meeting, April 22–24, 1987. (Washington: Library of Congress, 1987). 66 p. $7.50.

No. 15 *Nationwide Networking:* Proceedings of the Library of Congress Network Advisory Committee Meetings, July and December 1986. (Washington: Library of Congress, 1987). 65 p. $7.50.

No. 14 *The CONSER Project: Recommendations for the Future.* Report of a study conducted for the Library of Congress by Jeffrey Heynen and Julia C. Blixrud. (Washington: Library of Congress, 1986). 122 p. $7.50.

No. 13 *Toward a Common Vision in Library Networking:* Proceedings of the Library of Congress Network Advisory Committee Meeting, December 9–11, 1985. (Washington: Library of Congress, 1986). 95 p. $7.50.

No. 12 *Key Issues in the Networking Field Today:* Proceedings of the Library of Congress Network Advisory Committee Meeting, May 6–8, 1985. (Washington: Library of Congress, 1985). 88 p. $7.50.

No. 11 Maruyama, Lenore S. *The Library of Congress Network Advisory Committee, Its First Decade.* (Washington: Library of Congress, 1985). 50 p. $7.50.

No. 10 *The Information Economy in the U.S.; Its Effect on Libraries and Library Networks:* Proceedings of the Library of Congress Network Advisory Committee Meeting, November 14–16, 1984. (Washington: Library of Congress, 1985). 59 p. $7.50.

No. 9 *Electronic Information Delivery Systems:* Proceedings of the Library of Congress Network Advisory Committee Meeting, April 18–20, 1984. (Washington: Library of Congress, 1984). 79 p. $6.00.

No. 8 *Public/Private Sector Interactions: The Implications for Networking.* (Washington: Library of Congress, 1983). 48 p. $5.00.

No. 7 *Document Delivery—Background Papers Commissioned by the Network Advisory Committee.* (Washington: Library of Congress, 1982). various pagings. $5.00.

No. 6 Vondran, Raymond F. *National Union Catalog Experience: Implications for Network Planning.* (Washington: Library of Congress, 1980). 51 p. $4.00.

Continued on Back Cover

Network Planning Paper No. 19

BEYOND BIBLIOGRAPHIC DATA:

Proceedings of the
Library of Congress Network Advisory Committee Meeting
March 29-31, 1989

Network Development and MARC Standards Office
Library of Congress
Washington

1989

Library of Congress Cataloging-in-Publication Data

Network Advisory Committee. Meeting (1989 : Washington, D.C.)
Beyond bibliographic data : proceedings of the Library of Congress Network Advisory Committee meeting, March 29-31, 1989.

(Network Planning Paper ; no. 19)
Bibliography: p.
1. Library information networks--United States--Congresses. 2. Reference services (Libraries)--United States--Automation--Congresses. 3. Data bases--United States--Congresses.
I. Library of Congress. Network Development and MARC Standards Office. II. Title. III. Series.

Z674.8.N46 1989 021.6'5'0973-20 89-600331
ISBN 0-8444-0657-0

For sale by the Cataloging Distribution Service, Library of Congress, Washington, D.C., 20541.

CONTENTS

FOREWORD 1

ATTENDEES 3

INTRODUCTION, Sandra K. Paul 5

PROGRAM SESSION 9

USER PERSPECTIVES AND REQUIREMENTS: CREATOR OF NON-BIBLIOGRAPHIC DATABASES HAS TO SHARE WITH OTHERS, Rudolph M. Bell 9

MAKING REMOTELY SENSED DATA MORE ACCESSIBLE, Charles E. Olson Jr. 19

NETWORKING AND ACCESS TO NON-BIBLIOGRAPHIC DATABASES: A COMMERCIAL PERSPECTIVE, Dennis McDonald 25

RLG ASSESSMENT OF NON-BIBLIOGRAPHIC INFORMATION NEEDS, Constance Gould 31

SUMMARY OF THREE NON-BIBLIOGRAPHIC OPERATING ENTITIES, C. James Schmidt 39

BEYOND BIBLIOGRAPHY OR CREATING THE ROSETTA STONE FOR THE 21ST CENTURY, Kenneth B. Dowlin 43

NON-BIBLIOGRAPHIC DATABASES IN THE NETWORK CONTEXT: MEETING NOTES AND DISCUSSION COMMENTS, Ronald F. Miller 47

SUMMARY OF BUSINESS SESSION 53

APPENDICES

A. Meeting Agenda 55

B. Letter to Mr. Kastenmeier 57

FOREWORD

The extension of networks beyond bibliographic data is number twenty of an action agenda developed by the Library of Congress Network Advisory Committee in December 1986. The realization of the entire agenda will provide a clearer perception of a common vision in library networking.

The goals of this meeting were to gain a better understanding of the term "non-bibliographic" in the library network context, to gain a better understanding of the role of non-bibliographic databases in the library network environment, and to begin to appreciate the range and potential of such electronic information.

I gratefully acknowledge the assistance of the Program Subcommittee--Sandra K. Paul, SKP Associates, representing the Association of American Publishers; H. E. Broadbent, Pittsburgh Regional Library Center; James P. Riley, then at the Federal Library and Information Center, Library of Congress; and C. James Schmidt, The Research Libraries Group, (RLG) Inc.--in making the meeting a success. The Program Subcommittee joins me in thanking all those who gave presentations: Rudolph M. Bell, Rutgers University; Charles E. Olson, Jr., University of Michigan; Dennis McDonald, Online Computer Systems, Inc.; Forrest Williams, U.S. Bureau of the Census; Ward Shaw, Colorado Alliance for Research Libraries; Constance Gould and C. James Schmidt, RLG; Kenneth E. Dowlin, San Francisco Public Library; and Ronald F. Miller, Cooperative Library Agency for Systems and Services. Owing to time pressure Messrs. Shaw and Williams were unable to write a paper reflecting their presentation. The editorial work was done by Sigrid G. Harriman, Library of Congress.

The document has been issued within the Network Planning Paper series. It should be noted that the opinions expressed in the proceedings are those of the individual speakers and do not necessarily reflect the opinions of their organizations.

Henriette D. Avram
Chair, Network Advisory Committee

October 15, 1989

ATTENDEES

Organizations	Representatives
American Association of Law Libraries	Robert L. Oakley
American Library Association	Joseph F. Shubert Carol C. Henderson
American Society for Information Science	Ward Shaw
Association for Library and Information Science Education	Ann Woodsworth
Association of American Publishers	Sandra K. Paul
Association of Research Libraries	Prudence S. Adler
Bibliographical Center for Research	David H. Brunell
Capital Consortium Network	Dennis Reynolds
Chief Officers of State Library Agencies	Richard G. Akeroyd
Cooperative Library Agency for Systems and Services	Ronald F. Miller
Council on Library Resources	Deanna B. Marcum
Federal Library and Information Center Committee	D. Lee Power
Library of Congress	Henriette D. Avram (chair) Sigrid G. Harriman (secretariat)
Medical Library Association	Erika Love
Minnesota Interlibrary Telecommunications Exchange	Anita Anker Branin
National Commission on Libraries and Information Science	Susan K. Martin Sandra N. Milevski
National Agricultural Library	Pamela Q. Andre

National Library of Medicine	Lois Ann Colaianni
Nelinet, Inc.	Dale P. Flecker
OCLC Online Computer Library Center, Inc.	Mary Ellen Jacob
Pittsburgh Regional Library Center	H. E. Broadbent III
Southeastern Library Network	Frank P. Grisham
Society of American Archivists	Richard V. Szary
Special Libraries Association	Bette H. Dillehay
University of California	Dennis E. Smith
University of Chicago	Charles T. Payne
Utlas International	Richard W. Newman

Organizations	Observers
Library of Congress	Robert G. Zich

INTRODUCTION

In the past thirteen years, the Library of Congress Network Advisory Committee (NAC) focused most of its attention on the networking of bibliographic data. In recent years, NAC members have expressed concerns about non-bibliographic databases (full-text, numeric, graphic) being converted into machine-readable form and placed for inclusion in or already residing on information networks. There appeared to be very little effort underway to establish standards for these databases for efficient access and exchange.

The March 29-31, 1989, NAC meeting, held in Washington, D.C., was on the extension of networks beyond bibliographic data in the library network context.[1] The Planning Committee was chaired by Sandra K. Paul, president of SKP Associates, who represents the Association of American Publishers on NAC. Other members of the Committee were: Henriette D. Avram, Library of Congress and chair of NAC; H. E. Broadbent III, Pittsburgh Regional Library Center; James P. Riley, then at the Federal Library and Information Center, Library of Congress; and C. James Schmidt, Research Libraries Group, Inc.

In considering the nature of this topic, the Planning Committee decided that the meeting should be educational and provocative, but that it should also be realistic--including only examples of the type of information which is now or shortly will be made available through library networks. In addressing the scope of the topic, it was agreed that the coverage must be comprehensive--demonstrating a variety of the available full text, graphic, and numeric information. An attempt was made to include non-bibliographic information provided by individual scholars, societies and associations, commercial publishers and the government through local, regional, and national library networks. Finally the Planning Committee sought a keynote speaker whose futuristic views would challenge traditional assumptions of information availability through library networking.

After a welcome from Henriette Avram and brief introduction to

---

[1] "To evaluate the extension of networks beyond bibliographic data" was task 20 in the series of tasks identified as necessary steps toward the goal of the "common vision" at the December 1986 meeting. For a complete list of all tasks, see the proceedings of the July and December 1986 Library of Congress Network Advisory Committee meetings in Network Planning Paper No. 15, entitled *Nationwide Networking*. [Ed. note.]

the topic by Sandra Paul, presentations were made by seven individuals. Each presentation was followed by time for questions from NAC members and other speakers.

Representing the type of numeric, tabular, and text databases developed by the scholarly community, Rudolph L. Bell, professor of history at Rutgers University, described the Medieval and Early Modern Data Bank (MEMDB) of which he is co-director. MEMDB is composed of such information as price data, coin values, glossaries of weights and measures, and calendars of dates. Professor Bell discussed the problems of mounting and accessing such data, which lacks accepted coding and cataloging.

Charles E. Olson, Jr., professor of natural resources at the University of Michigan, addressed the electronic storage and access to photographic information assembled by a society. He reviewed the problems that can be solved by providing professional researchers with access to aerial photographs, maps, and graphs.

Dennis McDonald, vice president for information products at Online Computer Systems, represented the commercial sector. He reviewed the activities of his database and software publishing company in providing full text, numeric, and graphic data for their customers and considered the critical issues in accessing this type of data.

Forrest B. Williams, acting chief, Systems and Programming Branch of the Data Users Services Division of the U.S. Census Bureau, presented an overview of numeric data available from the government on CD-ROM.

Ward Shaw, executive director of the Colorado Alliance of Research Libraries (CARL), represented a regional network. He described the wide variety of non-bibliographic information available on CARL. CARL's database offerings include local information such as a public affairs listing for the city of Denver, as well as indexes to the journals subscribed to by CARL member libraries and summaries of the articles contained in those journals.

Constance Gould, program officer for program development at the Research Libraries Group, Inc. (RLG) presented the results of RLG's survey of twenty-five of its member institutions' assessment of the non-bibliographic information needs of academic faculty. She also read a paper prepared by C. James Schmidt, vice president and director of the Research Libraries Information Network (RLIN), presenting considerations for mounting non-bibliographic databases on a national research network.

The keynote address was the final presentation of the meeting. Kenneth Dowlin, director of the San Francisco Public Library, presented "Creating the Rosetta Stone for the 21st Century." He outlined his vision of tomorrow's library and made use of a video to show NAC members the variety of non-bibliographic information that the public will expect to find in the public library of the twenty-first century--only eleven years from now.

Ronald F. Miller, executive director of the Cooperative Library Agency for Systems and Services (CLASS), chaired the review and discussion period. After synthesizing all presentations he urged the attendees to comment on what they had learned during the program session. The growing demand for and problems in gaining access to the types of information described by the speakers were discussed. Mrs. Avram appointed a small committee[2] to explore the possibility of working with the U.S. National Commission on Libraries and Information Science (NCLIS) on an analysis of existing inventories of available non-bibliographic databases, as a first step in improving or systematizing access to both research and library networks.

Prepared by: Sandra K. Paul

---

[2] Members of the committee are Henriette D. Avram, chair; Susan K. Martin, the executive director of NCLIS, and Sandra Milevski from NCLIS. At this writing some background work has been done, both at NCLIS and at the Library of Congress. Discussions between Avram and Martin, considering steps for a funding proposal to identify the universe of electronic archives, followed. "To commission a study to collect existing inventories of electronic archives..." was task 19, identified by NAC and also issued in Network Planning Paper No. 15. [Ed. note.]

USER PERSPECTIVES AND REQUIREMENTS: Creator of Non-bibliographic Databases Has to Share with Others

Rudolph M. Bell[1]

The topic on which Henriette Avram asked me to speak today involves an injunction about which there is no room for disagreement. Of course, the creator of non-bibliographic databases has to share with others; the only question is how best to do so, or as I shall propose shortly, how better to do so. My vantage point on the problem is that of a professor of history, one who has created a few databases along the way and who uses these, along with others created by fellow scholars, both in teaching and in research. As co-director of The Medieval and Early Modern Data Bank (MEMDB), an electronic library of machine-readable information concerning Western history for the period circa A.D. 800-1800, I have been more involved with the issues raised by shared databases than the average historian, but the time already is at hand when all scholars will be affected by decisions reached and implemented at meetings such as this one.

Background

Let me recount briefly some of the circumstances encountered back in 1974 when I attempted to get a share of someone else's database. The material in question was cited in Robert Fogel and Stanley Engerman's supplementary volume to *Time on the Cross*[2], and consisted of some five thousand slave sales recorded in New Orleans, mostly in 1850s. Fogel and Engerman acknowledged that the compilation of their database had been funded in part by public granting agencies and that it was available to any interested scholar. Try as I might, however, all efforts at a bibliographic search failed to tell me where the database was, much less how to obtain a copy of it. A telephone

---

[1] Rudolph M. Bell is a professor in the Department of History at Rutgers State University of New Jersey. This paper is based on a presentation given on March 30, 1989, at the Library of Congress Network Advisory Committee meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of Rutgers University or its History Department. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

[2] *Time on the cross: the economics of American negro slavery*, by Robert William Fogel and Stanley L. Engerman. Lanham: University Press of America. c 1984, 286 p.

call to Stanley worked splendidly - he being alive and well, and cooperative - but this is hardly a model way to proceeding such matters.

In any event, the tape containing the database was registered in his name at the University of Rochester Computer Center and with his permission I obtained a copy. Since the data was heavily coded -a 17 in columns 34-35 meant 'unruly' whereas a 14 indicated a 'malingerer'- the tape was useless to me until Stanley also sent along a code book. This turned out to be a photocopy of a handwritten document of some 150 pages with various jottings here and there. I leave aside the many tribulations incurred before we actually got the tape up and running at Rutgers, and report only that when I published some results from my reanalysis of this data bank I was determined to do a better job of documentation than Fogel and Engerman had done. The effort was in vain. The Rutgers Library at that time had no means to catalog a computer tape and our Computer Center, like the one in Rochester, knew only how to file a tape under its owner's name, not according to its title or subject. The library did not deem either Stanley's handwritten code book or my Statistical Package for the Social Sciences (SPSS) printouts as analogous to a book, nor did they consider these loose sheets to be a manuscript, and so they could neither store nor catalog nor make known to others the existence of the work. To this day, if you want to know about New Orleans slave sales, the only way is to telephone Stanley or myself, or else to go back to the manuscript records themselves and start all over.

There must be a better way. Indeed, it is possible that there is a best way. However, I have serious reservations about the utility of striving to reach a 'best way' when it comes to non-bibliographic databases. The efforts made within the library community to achieve uniform standards for cataloging bibliographic information are entirely laudable, but they may not constitute an appropriate model for other kinds of data. Calls to develop a single international standard for the electronic representation, storage, and retrieval of information, even if limited to the discipline of history that I happen to know best, strike me as futile and even wrongheaded. Moreover, the very act of seeking such a lofty goal has the unintended consequence of undermining less ambitious but more practical efforts to achieve at least a better way.

Let us consider what was most absurd in the failures fifteen years ago of Fogel, Engerman, and then Bell to share a database. Firstly there was absolutely no reference to the database in any library; secondly, access to the database (one required by law because of its funding to be fully available to the public) required a secret password available only from the author; thirdly, use of anything in the database was possible only for someone comfortable with the jargon of tracks, bits-per-inch, record-lengths, and binary-coded-decimals; and fourthly, interpretation of the content of the database necessitated obtaining a copy of a detached, uncataloged, handwritten, crossed-over, and partially incorrect code book located in a cabinet in the author's office.

The Medieval and Early Modern Data Bank

MEMDB is designed to eliminate at least the worst of these problems. Therefore, MEMDB operates simultaneously at three distinct levels. First, it maintains a Master Data Set (MDS) that allows researchers to conduct interactive and integrated searches and retrievals of data stored online in a mainframe computer system operated by The Research Libraries Group, Inc. (RLG), in Mountain View, California. This retrieval system is accessible from any telephone line in the world connected to a Personal Computer (PC) using the Disk Operating System (DOS) and a modem. Second, MEMDB archives at its Rutgers University office individual data sets that are of too specialized a nature to justify inclusion in the master data set, and makes these archived data sets available on tape of floppy disk upon request. Third, MEMDB catalogues machine-readable data sets that are held by other archives and provides bibliographic information about these through RLG's Research Libraries Information Network (RLIN). Before untangling all of these acronyms and delving into each of the three levels of operation, you may wish to know a bit more about the history of MEMDB itself.

MEMDB was founded in 1982 at Rutgers University by my colleague Martha C. Howell and myself. Her research into the social and economic history of the late medieval and early modern Low Countries had been much hampered by the unavailability of coherent information about monetary values in that period, and convinced us of the need for the services MEMDB provides.

The incoherence of the sources that plagued Howell was a direct reflection of the incoherence of the monetary system itself, which of course was no system at all: not only kings, but also other autonomous powers, such as dukes, counts, city-states, and prince-bishops,could and did issue their own currencies in gold, silver, and copper. Most of these coins were then debased, clipped, and copied with such frequency as to make them unusable for long-term, and thus long-distance, trade. A few coins, such as the Florentine florin and the Venetian ducat, were sufficiently stable in value and widespread in distribution to be used as standards in international exchange, but contemporary official publications listing foreign coins in circulation and their value in the local currency were issued only infrequently.

Consequently, the historian confronting pre-modern wages and prices frequently has been at a loss. Faced, for example, with a price for cloth in Riga in 1356 expressed in marks of Luebeck, the historian has had to search for records giving the silver content of the Luebeck mark in 1356, and then has had either to locate an exchange between the mark and a currency of known value,or to estimate the mark's value by relating the commodity exchange rate implicit in the price to others.

Historians traditionally have dealt with this problem individually, each keeping separate records of exchange rates and sometimes publishing this information as appendices or footnotes to other studies, or, occasionally, in monograph form. It was the piecemeal nature of these publications that prompted us to conceive

originally of developing a computer-based listing of medieval currency exchanges or equivalent values, a listing that could be updated as necessary, searched, and delivered to users in some fashion.

The first major breakthrough came on the European side of our endeavor, when Peter Spufford of the University of Cambridge agreed to convert his work-in-progress on currency exchange quotations into a machine-readable format and then to donate his results to MEMDB. In 1986 his *Handbook of Medieval Exchange*[3] was published, and it is this fundamental work, listing over 13,000 exchanges dating from 1106 to 1509 and covering all of Europe, Byzantium, the Levant, and North Africa, that provided the first data set fully incorporated into MEMDB.

Shortly after our initial association with Spufford, Howell and I were joined by RLG. RLG, which already had more than a decade of experience in managing the standardization, acquisition, maintenance, and distribution of bibliographic data, was interested in expanding its scope to include non-bibliographic information, and arranged to cosponsor MEMDB. Spufford had offered an outstanding example of the kind of data appropriate for inclusion in our Bank, and now RLG provided a sophisticated and thoroughly tested means of dissemination.

Proposed Development of MEMDB

In December 1985 a group of medievalists from the United States, Canada, and The Netherlands met with us and with RLG planners to help draft the outlines of MEMDB's proposed development. From this meeting there emerged a permanent Advisory Board, whose members now represent leading American, Canadian, and European universities.

Our next major step involved the decision to develop a PC-based prototype of the master data set using Spufford's currency exchange quotations as the database. Rather than leaving the design and implementation of the prototype entirely to the experts at RLG, we involved ourselves at every stage. The moments of frustration and misunderstanding on both sides were many, but we are convinced, as are the system analysts and programmers at RLG, that without them the prototype would not have been possible, while without our input it would have been less good than we believe it is. The prototype is available at a nominal charge (to cover the costs of commercial software packages included within it).

The prototype embodies several central MEMDB principles concerning the ways in which machine-readable data sets should be made accessible. Firstly, the system had to retain all the documentation conveyed with the data set, not just the tabular results, and had to make this documentation easily available. Scholars, familiar with Spufford's book, or with others of this genre, already know that the

---

[3] *Handbook of medieval exchange*, by Peter Spufford with the assistance of Wendy Wilkinson and Sarah Tolley. London: Offices of the Royal Historical Society, 1986.

documentation consists not only of source references, but also of explanatory texts that are essential for understanding the data entries. In the case of Spufford's *Handbook*, which includes a lengthy introduction that is necessary for the specialist and the neophyte alike, we concluded that a user of MEMDB who wished to retrieve simply a single currency exchange quotation nevertheless should be able easily to consult online all potentially useful explanatory material, starting with commentary relevant to the specific quotation but including even the full introduction. This decision gave RLG specialists some interesting challenges, but they rose to the occasion. The prototype does, and the full RLIN system will, provide complete textual support for all its data sets, with correct linkages provided for the user of specific text to specific tabular data.

Secondly, the system had to be extremely user-friendly, neither mystifying and frustrating to the novice who is unfamiliar with computers and may require frequent prompting or assistance, nor cumbersome and annoying to the experienced user who wants to be able to get results with as little interference as possible. The prototype uses only eight command verbs, eleven index nouns, and a minimum of qualifiers (and, or, not) and punctuation ("","",-,/) to handle even the most complex searches of the Spufford data set. The structure, or syntax, of the command language also is easy to learn, and in fact many early users have simply figured out on their own how it works without any need to consult the manual. While some expansion of the vocabulary and syntax will become necessary as different kinds of materials are added to the master data set, we will remain committed to the goal of making MEMDB accessible to all scholars, regardless of their level of computer expertise.

Thirdly, the system had to be accessible at virtually all times, at a reasonable cost, and without the need for sophisticated personal computers. Development of the PC-prototype convinced us that only a design involving both PC's and a mainframe would work.

This integrated design differs from other data banks in several important ways. Most are offline only, such as the holdings of the Inter-university Consortium for Political and Social Research (ICPSR), the Cambridge Group for the History of Population and Social Structure, or the Danish Data Archives. Others are bibliographies or catalogues, such as the Fontes Anglo-Saxonici at the Universities of Cambridge and Manchester, England. A third type of data bank supplies textual material, such as the Thesaurus Linguae Graecae project or the American and French Research on the Treasury of the French Language (ARTFL), but these typically require a dedicated computer. A final category is represented by the Max Planck Institute's "Kleio" project, which is developing software for creating and retrieving machine-readable historical data sets at a sophisticated work station equipped with an optical laser disk reader and a very large hard disk. The possibilities offered by this venture are intriguing, but would require individual scholars to make a considerable investment in equipment (about four to eight times the cost of the simple PC needed to access MEMDB on RLIN). Moreover, use of the system involves a level of competence in computer usage that remains uncommon even among recently-

trained historians. Nor does current technology allow for easy and uniform updating of data sets in the system.

Distribution of MEMDB

In 1990 MEMDB will migrate from its PC-prototype to distribution via RLIN. In North American, access to MEMDB through any of the more than 1500 existing RLIN terminals, or for that matter via telephone line by dialing a local TELENet number, is not a problem. For Europe the situation is more complex, but now we are reaching a solution.

While exact pricing remains to be determined and obviously is subject to change, a charge of less than $1 per completed query is anticipated, to which must be added the local telephone charges. A completed query, of course, might result in only a few data items, but it may also provide the user with large amounts of information culled from a variety of discrete holdings within the master data set, all available within minutes, if not within seconds.

You may ask just what will be available in MEMDB's Master Data Set, and why the master data set concept is important. The master data set will include all works of a primarily tabular nature concerning the medieval and early modern history of the West that one would expect to find in the reference or consultation section of any good university library. It will also include bibliographic references to machine-readable data sets not incorporated in the master data set itself and, in the future, glossaries of weights and measures, gazetteers of Latin and vernacular place names, and calendars of dates. Slated for inclusion in the master data set before it becomes available via RLIN in 1990 are data sets on medieval mint production, a Rome census of 1526, some 1,200 currency exchange quotations donated by Spufford as an update to his original data set, and medieval and early modern price data from Germany and The Netherlands.

This is an eclectic list, the result of our decision to work out certain technical design questions before imposing a more traditional historical scheme upon which to base the expansion of the master data set, and even from this short list we perhaps may draw an example to illustrate the advantages of a truly integrated master data set. A synthesis of currency exchanges and price data would be an arduous task using traditional printed sources, and still very difficult treating them as two discrete machine-readable data sets, doing separate calculations of averages and trend lines for each set, adjusting for their differing variable structures, and keeping track of cautionary notations, exceptions, and qualifications. In contrast, once the two sets are in the master data set, the researcher has instant access to facilities for combining data entries from either set, or for separating them when necessary, whether by date, by type of commodity, by author, by price range, etc. Add the mint data and silver basis may be constructed across different currencies; eventually, add a perpetual calendar and the user may quickly determine whether prices were higher just before Easter.

The Master Data Set

In a more imaginative vein, one may think of the master data set as placing the scholar in a room lined with the standard reference works, hundreds or even thousands of them. Via RLIN the user consulting MEMDB employs just a dozen or so commands to open instantly any or all volumes, to consult any or all of their tens of thousands of pages of information simultaneously, to view the full documentation for each individual data entry at a single keystroke, and to construct and retrieve in electronic form for further analysis an essentially new data result set culled from tens or even hundreds of different works. Yes, dozens of persons working for hundreds of hours could accomplish the same thing, and the scholar's eye for what is crucial remains indispensable, but surely the technological time is right for using machines to expand vastly the researcher's vista. This is what the master data set accomplishes. Again, an analogy with more traditional library science considerations may be helpful. The most frequently consulted reference works in any library properly are placed on open shelves in a section where users may browse at will and do not need to fill out slips to call up a particular volume. It is this sort of work that belongs in the master data set.

Building the full master data set will take time. Most critical is the acquisition of new work being done by scholars. The vast scope of MEMDB does not allow us to support directly the archival work of individual scholars. Our task, instead, is the fullest and most effective integration and dissemination of such work. To this end, our offices in Europe and at Rutgers are ready to provide advice and consultation at every stage of prospective research activity: beginning with queries about existing and possible related data sets, continuing with suggestions on appropriate coding strategies and software packages, and ending with publication (in machine-readable form) of completed results. In theory MEMDB can work with virtually any machine-readable format, but in practice some ways of doing things are far better than other ways, both for the originating scholar and for the editors at MEMDB. Analytic texts based on such work surely will continue to appear in print media, but the raw data, now so frequently either unavailable or available only in cumbersome microfiche or sui generis private machine-readable forms, should find an accessible, integrated, and permanent place of electronic publication via deposit at MEMDB.

Then there is the vast array of works already in print, many of them done laboriously by nineteenth- and early twentieth-century scholars, upon which medieval and early modern scholars depend. These, too, belong in a good research library, and therefore in MEMDB. We use both optical scanning and keying to retrieve such materials, and of course first obtain copyright clearance where that is necessary. The problem here is one of money, although as library specialists become increasingly aware of the enormous costs of maintaining and restoring rapidly deteriorating volumes printed on acidic paper, it becomes evident that an alternative such as MEMDB, with its full retention of all documentation along with tabular material, is an option well worth supporting.

MEMDB's Offline Data Set Archive

The second level at which MEMDB operates, its offline data set archive, is if no less significance than the master data set. As with data sets in the master data set, holdings of the offline archive constitute published works made available to researchers in electronic form. So also, bibliographic references to the archived data sets are fully integrated within the RLIN master file and may be accessed via author, title, publisher, and all the other categories used within RLIN.

To continue with the library analogy, other valuable but more specialized and therefore less frequently consulted volumes in a traditional library are reached via a catalog reference and then a call slip. The offline archive operates in a similar fashion, with due allowance for the differences entailed in a heavily computer-based operation. The user starts with RLIN and the master data set and locates the reference to a data set held in the offline archive. Without leaving the RLIN environment, the user may request access to an offline data set; after specifying the mode in which the data are to be sent (tape, diskette, printout, etc.), the user's request is transmitted from RLIN to the offline archive, where the request is translated by an MEMDB staff member into a batch-processed job and prepared for delivery, either by standard mail service or electronically depending on the user's specifications.

In theory every data set could be part of the master data set, just as in theory a library could keep all its shelves open, but the negative aspects in both traditional and electronic libraries outweigh the advantages of instant access to more specialized items. One of the major reasons for the master data set is its facility for integrating searches for related information across different data sets. Therefore, works that do not integrate well with the general character of master data set holdings, for example the individual baptismal records of a single parish or the shipments of a particular merchant, belong more properly in the offline archive. To include all data sets in the master data set, apart from the cost considerations involved, would result in overwhelming the user with data that in most cases would turn out to be an irrelevant distraction.

Another reason for distinguishing between master data set holdings and those held in archive is that data sets in the master file aim for individual variables that are self-explanatory and as close as possible to the raw data, whereas the offline archive is in a better position to handle and provide proper documentation for data that has been re-coded, grouped, averaged, or in other ways transformed. Finally, any data set donated to MEMDB with access restrictions (something we advise strongly against, but realistically cannot exclude) can only be made available through the offline archive. Again to invoke the traditional library analogy, works on the open reference shelves may be looked at and presumably copied by anyone; no one knows who has looked at what. Once a call slip is needed to access a closed-shelf work, restrictions and controls become possible; and a record may

be kept of who has consulted what. So also with the RLIN/MDS versus the offline archive division of data sets.

We would not wish to set forth a list of hard-and-fast rules about exactly what data sets go where, especially since the electronic library is so new and the decisions we make are not irreversible. However, some important distinctions between the handling of data sets in the online versus offline environments seem worthy of note. Although we try to protect the integrity of data sets in the master data set, it would be absurd to edit them to appear exactly as they came to us in their original form. To take only the most obvious, in editing Spufford's *Handbook*, we cleared up a few spelling inconsistencies, substituted paged cross-references in the original with pointers to content now residing in an electronic and unpaginated data set, standardized singular versus plural usages, added a decimal-equivalent field, and so forth. Moreover, in editing the background text, we had to make firm decisions about what verbiage would be linked to precisely which currency quotations, and that cannot be exactly the same as the experience of the reader whose eye may scan upward or downward and whose finger may flip pages. For these and other reasons, all data sets included in the master data set also are available in the exact form in which they ere given to use as part of the offline archive holdings. When a user retrieves and keeps electronically a result set drawn from the master data set, it is provided in a standard RLIN format, without integrated documentation (but with keys to both background text and to sources).

Data sets held only in the offline archive, on the other hand, will not be edited by us in any way other than to impose a consistency in external file and tape labels. Users who request one of these data sets will receive it in exactly the form in which it was deposited, including the particulars of the software used. What we received in Lotus will go out in Lotus, and so also for SPSS, dBASE III, Paradox, Word Perfect, or any other software package. Source and background information also will go out exactly as it came in, whether embedded in tabular data or appended as separate files.

Another key difference is that the master data set is constantly updated with new and integrated data. For example, to Spufford's original published *Handbook* quotations we are now adding another 1,200 or so that he received too late for checking and inclusion. These will be marked as a separate contribution and source, but in all other ways they will be fully integrated with the earlier data. Were both of these data sets only in the offline archive, the user would have to request both of them, receive them as separate files (perhaps even in differing formats), and then work out any necessary integration of the two sets.

MEMDB's Use of Electronic Holdings of Other Data Archives

The third level of MEMDB operations rounds out the first two, and for the last time in this talk I again invoke the analogy of the traditional library. In addition to standard reference works the open shelves and catalog cards leading to more specialized works, . good

reference library allows the scholar to consult the holdings of other libraries. So also, MEMDB through RLIN allows the user to consult the electronic holdings of other data archives.

All bibliographic data that come to MEMDB will be put into the international machine-readable cataloging (MARC) format specified for machine-readable data files. These entries will be accessible directly in the wider RLIN environment and to MEMDB/MDS online users via an easy access software switching facility. For example, a researcher exploring currency exchange quotations held in MEMDB/MDS will find many that come from Professor John Munro of the University of Toronto through Spufford's *Handbook*; the user may then search the master data set for other contributions by Munro, perhaps on cloth prices, and these too the user may request. Finally, the user could search RLIN's machine-readable data file to locate references to yet other data sets from Munro (or on cloth prices) held by other archives. Resulting bibliographic records may be requested in "long," "full," or "partial" versions depending upon the user's needs.

This, then, constitutes a brief introduction to the Medieval and Early Modern Data Bank as it now exists and as it plans to grow in Europe and in North American. I began by indicating some of the shortcomings in my early efforts to share in and then share with others a database on slave sales in New Orleans. To not mince words, that database was and is unknown, unknowable, unlisted, unavailable, and incomprehensible. MEMDB offers a way, not necessarily the only way, to make information available more readily.

MAKING REMOTELY SENSED DATA MORE ACCESSIBLE[1]

Charles E. Olson, Jr.[2]

I am an image interpreter. My specialty is identifying land cover, and through it land use, from aerial photographs and other remotely sensed data. This often involves determination of land use at some date in the past, or reconstructing the pattern of land use change over a period of years. My education as a forester, my military training in urban area analysis from aerial photographs, and my employment as an Extension Forester in the Cooperative Agricultural Extension Service in Illinois provide important building blocks for my work. Even this diversity of experience is inadequate to meet all of the challenges encountered in many situations. Some of those situations illustrate the importance of both textual and non-textual databases in solving problems related to land use.

Case I

Six years ago I was asked for help in establishing land use along a portion of the right-of-way for a gas pipeline. The pipeline crossed a low area and a new owner decided to convert the low area to a lake and subdivide the surrounding area as lake-front lots. The pipeline company sued to prevent lake construction on the grounds that a lake would effectively deny access to the easement (the pipeline). The property owner claimed the low area had always had water in it, and changing the depth of the water was not a major alteration affecting the terms of the easement. At issue were the claim that the area had always had water in it, and two specific aerial photographs showing the

---

[1] The original presentation was illustrated with several black-and-white and color images which are not included in this printed version.

[2] Charles E. Olson Jr. is a professor of natural resources at the University of Michigan in Ann Arbor, Michigan. This paper is based on a presentation given on March 30, 1989, at the Library of Congress Network Advisory Committee meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of the University of Michigan or its Department of Natural Resources. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

low area with substantial amounts of standing water.

Records from the National Cartographic Information Center (NCIC) revealed a long series of aerial photographs of the area. Examination of the October 1949 photograph revealed corn being harvested from the low area at the time the photographs were being taken. The 1955 photographs showed a linear disturbance in the low area which proved to be the tracks made by one land owner building a fence to keep his neighbor's sheep out of his hay field. These two sets of photographs showed the area had not always had water in it, but there were still the 1971 and 1972 photographs showing much of the area covered with standing water. Both sets of photographs had been taken in early April, and this timing proved significant. A check of weather records in *Climatologic Data for Michigan* revealed the photographs had been taken shortly after a warm period in late March during which more than 1.2 inches of rain had fallen. The heavy rain, coupled with melting snow, seemed to be a possible cause of the standing water in the low area. A further check of soil temperature records kept by Michigan State University revealed soil temperature on a gentle, south-facing slope were still twenty-nine degrees Fahrenheit, twenty-four inches below the surface, on the day the photographs had been taken. Low areas are frost pockets--cold air collects in them at night. This made it highly likely ground beneath the low area was frozen, and impervious, when the photographs were taken, providing perfect conditions for ponding of run-off and melt water.

Both the historic records--the old aerial photographs representing several dates--and the textual data in the weather records were essential to these determinations. Without the NCIC index, identifying the existence and sources of all the photographs required to meet demands of this case would have been at least difficult.

Case II

More recently, it was necessary to establish the location of the ordinary high water line (OHWL) along a stretch of river which had been strip-mined. The deferred royalties which were the basis of the litigation could reach $10,000 per acre, and a relatively small change in the location of the OHWL would mean a significant change in the settlement. Temporally, the best record of conditions before the strip-mining occurred was found in some 1940 aerial photographs held at the National Archives, but these were not of the best quality. Again, a search of the NCIC records led us to a series of photographs taken at different seasons of different years. One set of photographs showed standing water on the river flood plain. A check of hydrologic records revealed the photographs had been taken shortly after the river flood had crested within a few inches of the historic ordinary high water level. Comparative interpretation of the 1940 and later photographs let us transfer the OHWL to the 1940 photographs in those areas which had not been strip-mined in the later photos. With these segments of the OHWL identified, it was relatively easy to extend these line segments and identify the OHWL along the entire river valley. Again, a successful solution required merging image and textual data sources which would have been very difficult to locate without accessible

archival records.

Case III

In 1973, members of the National Academy of Sciences Committee on Remote Sensing Programs for Earth Resource Surveys (CORSPERS) were asked to evaluate early research results with data from the Earth Resources Technology Satellite (ERTS-1, later named Landsat). As the chair of CORSPERS' Biology Panel, I reviewed a number of projects related to forest and range resources. During a break in the ERTS Investigator's Symposium, one of the range specialists told me of a strange streak across his ERTS image of part of Wyoming. The streak was between one and five miles wide, trended WNW (West-North-West) to ESE (East-South-East), and was visible across the full one hundred mile width of the image. Careful checking of the data showed the streak to be real, neither an artifact of the data recording system nor a thin band of high cirrus clouds. The research team had given up trying to explain it when a colleague in the History Department happened to hear about the mysterious streak. To him, its interpretation was obvious--it was a remnant of the Oregon Trail, detectable from space more the one hundred years after its heyday.

In this case, the outside information needed to complete the interpretation of the image came from another person, rather than an archival record. While the archival record exists, the image interpreters had no idea where to look. Sharing such information is the essence of the "old boy networks" that often evolve among research workers in any field. The newcomer who is not privy to these networks needs another source.

Case IV

In 1979, the Environmental Protection Agency (EPA) was looking for a way to rapidly assess environmental impacts of proposed coal mining in the western United States. While working on this project, we located original aerial photographs in Bureau of Land Management (BLM) field offices which were not registered in any index. Such locally obtained photographs can provide important data sources for several types of land use studies, but are at least difficult to locate without direct contact with each local field office.

BLM is not the only Federal agency with nearly autonomous regional or district offices; offices which often compete with each other within an agency. Sharing data between such offices is seldom standard operating procedure, and even "headquarters" does not know what is held by its subordinate field offices.

Direct contact with all of the field offices of each agency can be time consuming, and the efficacy of these contacts is often dependent upon the person, or persons, with whom such contact is made--especially when the important photographs are the old ones which are often considered obsolete. Even when the contact person(s) know what photographs are held at the office, time required to locate all data sources essentially precludes the rapid response desired by EPA.

Case V

Sharing of data between agencies is fraught with turf-protection pitfalls. What records do exist are in different formats in different agencies. There is no single query which can locate data which already exists and is potentially available. An example from overseas which does not reflect on any U.S. agency may be useful.

The Environmental Research Institute of Michigan (ERIM) was asked to prepare a soils map of a country in Africa. The project was completed and the final map turned over to the Ministry of Soils. Later, ERIM was asked to return to the same country and complete a forest inventory. A first step in this new project was to ask the Ministry of Soils for a copy of the soils map prepared two years earlier. The response was, essentially: *This is ours, go make your own.*

Case VI

Another case involved a new township zoning ordinance and an existing non-conforming use. Like most zoning ordinances, this one was enacted with a grandfather clause which exempted existing non-conforming uses if they did not expand. Several non-conformers did expand and the township sued to reverse the expansion. The critical issue was determining the extent of non-conforming use on the effective date of the ordinance. There were no field records, and no aerial photographs had been taken of the area on the effective date. The best the township could hope to find were two sets of photographs bracketing the effective date as closely as possible.

A search of the NCIC led us to photographs taken ten months before and fourteen months after the effective date of the ordinance. Searching different sources not included in the NCIC index provided what was needed. Research to develop better remote sensors had resulted in considerable airborne data acquisition including the area in question. Few of these photographs are recorded at NCIC. Using the "old boy network," a set of photographs taken just days after the effective date of the ordinance were located. With these, we were able to document the extent of the non-conforming uses before, and shortly after, the effective date of the ordinance, and substantially later when the non-conforming uses had greatly expanded.

Final Remarks

The value of the many image resources held by many agencies and organizations has been repeatedly proven. The ability to locate these resources gives me, and others like me, a particular value to my students and clients. To make these resources more widely accessible requires something better that the "old boy network" now fragmentally in place. It may also require some changes in indexing procedures.

Each of the situations I have described required merging data from more than one source. Most required locating aerial photographs

from different dates and several sources. In addition to aerial photographs, many forms of maps, satellite images, ground photographs, and sketches are important sources of graphic information. These records are especially valuable when accompanied by a description of how and why the records were compiled, information which usually accompanies maps and photographs in research reports/monographs. Unfortunately, traditional cataloging procedures usually dispose of such records with a cryptic "2 maps," or "2 plates," added to the entry. Such entries are not very helpful without some information on scale, area covered, information portrayed, and the source of original data.

Because of the investment many agencies have in their own indexing systems, they are reluctant to change to someone else's system. An independent organization with no vested interest in any existing system may be able to access and convert records in many agencies to a single format. I believe the plan outlined by the Research Libraries Group (RLG) is workable, and RLG appears to be a logical organization to undertake the task.

NETWORKING AND ACCESS TO NON-BIBLIOGRAPHIC DATABASES: A Commercial Perspective

Dennis McDonald[1]

Online Computer System's Perspective[2]

Online has its feet in several camps. We build databases, we publish software, and we provide services for libraries, publishers, corporations, and others.

As software publisher, for example, we publish Opti-Net™, a product which enables a NETBIOS[3]-compatible local area network to access a Compact Disk-Read Only Memory (CD-ROM) database from the various nodes in the network. This is useful, for example when a library wants to make a product such as *The Electronic Encyclopedia from Grolier* on CD-ROM available from a variety of workstations.

In the library marketplace, we work for organizations such as the Library of Congress, the U.S. Naval Research Laboratory, and The British Library. Our work in this sector includes both mainframe and personal computer (PC) applications on CD-ROM as well as WORM[4] systems running under a computer operating system called UNIX.[5]

For publishers we produce optical disk versions of products as

---

[1] Dennis McDonald, PhD., is vice president for information products at Online Computer Systems in Germantown, Maryland. This paper is based on a presentation given on March 30, 1989, at the Library of Congress Network Advisory Committee meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of Online Computer Systems. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

[2] Online Computer Systems Inc. is hereinafter cited as "Online."

[3] NETBIOS is an IBM developed software interface between a network and networking software.

[4] Write Once Read Many, a system for writing digital data directly to an optical disk without the need for a separate mastering and disk stamping device.

[5] UNIX is a trademark of the AT&T Bell Laboratories.

The Electronic Encyclopedia from Grolier and Computer Aided Product Selection (CAPS™) system for Cahners.[6] For our clients, we publish text-only databases as well as databases that combine text with color or black and white images. Some of our optical disk products are already available in print or online form, while others are being published for the first time as optical disk products.

Definition of the Term "Non-Bibliographic"

I include three kinds of data under the non-bibliographic label--full text, numeric, and graphic. It's a simple way to divide up the world, and it will serve my purposes today. Each data type provides particular challenges to delivery via networks, and I should add, to deliver via other media as well, including optical media, which is one of Online's specialties.

I seriously question the term "non-bibliographic" data; calling a body of data "non-bibliographic" implies that it "lacks" something, when actually you could say that the exact opposite is true, that bibliographic data leads one to find full-text, numeric, or graphic information.

The reality is that there is no perfect medium available yet that combines the ideal features of infinite storage capacity, zero access time, and compatibility with all available hardware and software systems. Everything involves a tradeoff of some sort, and that includes use of networks for the delivery of commercially produced non-bibliographic data.

Is Demand for Non-Bibliographic Data Growing

Two important questions are, one, is the demand for non-bibliographic information growing and two, how is this affecting networking? There is certain circumstantial evidence that the demand is growing. The kinds of databases people approach Online about are increasingly of the non-bibliographic variety, particularly in terms of graphics handling. Industry-wide, the trend in delivery vehicles does appear to be going in the non-bibliographic direction. Users are becoming accustomed to graphic interfaces, and IBM class personal computers are becoming better at handling graphic applications as the move towards a Macintosh-like interface gains momentum.

Spreadsheets and statistical programs now are available on the PC that share many features with their mainframe brethren. Even "text retrieval" programs are coming down to the PC level either for access to fixed media such as CD-ROMs or as serious tools for managing document collections already in machine-readable form. All of this, I am certain, must eventually filter back to commercial and library networks.

---

[6] Cahners Technical Information Service in Newton, Massachusetts, publisher of the CAPS product, for which Online Computer Systems Inc. developed the retrieval software.

Challenges with the Handling of Non-Bibliographic Data

There are some real challenges with the handling of non-bibliographic data by networks (and with optical media as well). First of all, full text "..ain't what it used to be." What it used to be, and what it still is in many cases, is the availability online of indexed American Standards Code for Information Interchange (ASCII) text files of already published literature. Now full text publications are being transferred to optical media, a good case in point being *The Electronic Encyclopedia from Grolier*.

Also significant is that, with the current upsurge in demand for graphic interfaces and with the proliferation of laser printing and desktop publishing, at Online we are seeing demand for going beyond pure text files to either the calling up of referenced color or black and white images from within a full-text file, or even the retrieval, recomposition, and display of print-quality images at a user's workstation.

One way to do this, of course, is to scan and present raster-scanned images directly to the user; this has been one approach we have used to present previously published documents to users. Another approach taken by some is to re-compose and re-create documents "on the fly" so that what the user at the retrieval work station sees is what looks like a page with all its typographic qualities--even if the document hasn't appeared on paper before.

Both approaches are potentially expensive. There is the issue of monitor and display quality to deal with; a low-end, character-oriented system without graphics capabilities isn't appropriate, say, when you are trying to display four point type on an engineering drawing. Significant data preparation is also necessary to tag the source data so that it will format correctly on screen and link embedded references within the text to appropriate scanned images. Who pays for this? And who pays if hypertext linking among documents must be built in when the document is being prepared? The user, obviously, in a commercial situation, although there is a market for high-end systems that mimic the quality and appearance of paper on screen. How large this market is, though, I can't really say, and this isn't even taking into account rising demand for the storage and retrieval of CAD[7] images via local networks in a corporate setting.

Displaying a high quality image sometimes requires users to sit through hardware or software decompression time, and this may only occur after all the information has been downloaded from a host or local optical medium. Will users wait in order to obtain the high quality image in a commercial situation? And what do you enable the user to do with the image after it appears on the screen? (I'm not

---

[7] Computer Aided Design; a system to be used at workstations by engineers and architects to develop two and three dimensional drawings of objects.

even going to get into the special problems of displaying quality color!) As I said, full text "...ain't what it used to be."

Numeric Data

It may seem like a truism, but numeric data needs to be analyzed before it becomes usable. Question is, where do you put the processing power--on the network, or at the user's site?

Different people solve this in different ways. One way is to sell discrete "time series" of data, data that have already been put into a meaningful order or arrangement and which are ready to be analyzed. Optical media can also be used in a similar fashion. Where data sets are accessed and downloaded from a mainframe, groups or packets of numeric data can also be downloaded from an optical disk, and depending on the amount of data being read, the reading of numeric data from an optical disk can also be time consuming.

There is a very strong commercial incentive to segment data for commercial sale, e.g., by identifying what I call the "minimum sellable unit" that is meaningful to the client. This is one approach taken, for example, by vendors of marketing data.

One potential difficulty of offering numeric data on a commercial network is that there are so many discrete bodies of data and user groups. This raises the problem of how many large numeric data collections can be offered profitably via a network. For small specialized data sets it may be that distribution on optical media is more cost-effective. Besides, you need to take into account what people want to do with the numeric data they obtain via a network. For example, it may not make much sense to pay network and mainframe use rates if you can perform an exploratory analysis of a discrete data set using a sophisticated PC or mini statistical package.

This returns us to the "division of labor" between the individual workstation and the network. What I think we may be seeing is a move to provide more network-specific software at the user level, i.e., software designed to provide communications capabilities and, beyond that, retrieval and analytical capabilities at the workstation.

What do you Mean, Only Six Hundred Megabytes

I am constantly reminded these days of the limitations of a CD-ROM disk; this is nowhere more apparent than with databases of graphics only or of text with graphics.

Where it may take many minutes to download a drawing or image from a network, it may take only a few seconds to take an image off CD-ROM and decompress it for display. However, six hundred megabytes can fill up fairly rapidly with image data, particularly if you want color, which takes a lot of space. Nevertheless, the demand for color is increasing. Applications which may formerly have gone to video disk and its fifty-four thousand-image per side capacity are now looking at CD-ROM and its potential for displaying high quality digital images

using relatively inexpensive off-the shelf PC hardware.

This also brings up an issue that is relevant to both networks and local optical storage--just getting the image to the user isn't enough. It has to be displayed and sometimes printed. There are a variety of processing and display formats available, with some VGA[8] modes and image processing algorithms providing particularly good results. Still, low-end color printing is still in its infancy, with printers still expensive. And looking at the other end, scanners to get color images digitized cost upwards of $5,000-7,000, though these prices are bound to come down soon, too.

Responding to Demand

Taking some of these comments to heart, what are some of the things networks can do to increase access to non-bibliographic data files?

The first thing is to store files locally that can't be economically accessed online. Networks can provide users the ability to switch back and forth between local optical databases for, say image data and for relatively static files that need to be accessed often, with the host reserved for recent information and heavy duty data processing.

Another approach is to make the application software that accesses the network more powerful or user friendly, perhaps by providing a graphical front end that makes it easier to steer the user towards certain commands or user features. Some commercial networks already provide front-ends like this, for example, some investment packages, and graphic interfaces such as Compuserve's Navigator and MacNet's graphic interface.

Another approach is to make it easier to integrate network-accessed data with data available locally. The simplest examples are providing numeric data in downloadable files compatible with standard packages such as PC spreadsheet packages.

Conclusions and Predictions

First, I think we will see the development of hybrid systems that use local optical storage for certain files while still linking with the host. I don't just mean telecommunications software to switch back and forth between the two, I mean real database access systems where there is a true integration of the online and the local data. Pressure for this comes from people in the commercial sector requiring updates and the need to access data files that can't be quickly or economically accessed online. (I personally think that the need for regular updates in some market sectors is overblown, but that's another story entirely.)

---

[8] Video Graphics Adapter, a computer display technical standard.

Second, local workstations will become more powerful. By "powerful" I mean that more of the user interface software will move to the user's workstation, making possible individual activities that might not be economically performed at the host level. A graphical interface is the most obvious of these. In other words, we may move even farther away from a common user interface than we already are.

Finally, blurring of local versus network access will prove to be a marketing challenge--and expensive. It's going to be a marketing challenge since developing hardware and software configuration that is appropriate for more than one particular application or family of databases may require a lot of planning and research before an appropriate market segment is identified. We may find that such integrated systems must be tailored very specifically to the application, thus making them commercially suitable only for relatively large (or rich) market segments than can afford them.

On the other hand, hybrid systems that link local and network-accessed databases may actually open up new markets, since it may be possible to combine access to a variety of different data files through a common user interface that is developed specifically to support a limited number of related data files and formats. Weaving together (and simplifying) access to more than one related file may well be viewed as a significant user benefit...and one which is potentially commercially successful.

THE RESEARCH LIBRARIES GROUP
ASSESSMENT OF NON-BIBLIOGRAPHIC INFORMATION NEEDS

Constance Gould[1]

Some of you may have wondered how and why the Research Libraries Group (RLG) decided to assess what kinds of information scholars need. Let me give you some background. Several years ago, RLG saw an opportunity to expand its services beyond the delivery of bibliographic information. The time seemed right: personal computers (PC) had begun to proliferate on member campuses, and we sensed a demand for new types of information in machine-readable form. We knew we had the skills and the computer network to deliver the information. But was it a pipe dream--or was there a real need?

We wanted to know whether the delivery of non-bibliographic information would be useful; and if so, what information was most needed. We asked ourselves how we could answer these questions, and it seemed obvious that the best way was to ask the people that we thought might need it--scholars and researchers. So the survey idea was developed and initiated.

We interviewed more than 250 faculty in the humanities, sciences, and social sciences at twenty-five of our member institutions. Essentially, we asked them four open-ended questions:

(1) How has your discipline changed in the last fifteen to twenty years; what are its new frontiers?

(2) do the changes imply a need to exploit data resources in new ways, or to make new information available?

(3) what types of information do you use in research: published material, all types of unpublished material (numeric, visual, archival, research in progress)? and

---

[1] Constance Gould, PhD., is program officer for program development at The Research Libraries Group, Inc. (RLG) in Mountain View, California. This paper is based on a presentation given on March 30, 1989, at the Library of Congress Network Advisory Committee meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of RLG. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

(4) what do you have good access to, and what do you need better access to?

Scholars confirmed our hunch that there is a need for the delivery of non-bibliographic information, and they told us where the needs are greatest. Many of you have seen the humanities assessment, which RLG published in September 1988, so you know what the humanities scholars said. What about the others?

Scientists and social scientists identified repeatedly two major problem areas: research in progress, and computer files. In this paper I will tell you the answers scholars gave: what is the scope of the problem in research in progress and computer files, what we can do about it, and what the benefits will be if we solve the problem.

RRESEARCH IN PROGRESS

Our question was: how necessary is it for scholars in the sciences and social sciences to have quick access to research in progress? We thought that in some fields it was more necessary than in others. Most people recognize that scholars in fast moving fields like superconductivity or Acquired Immune Deficiency Syndrome (AIDS) research cannot rely on what is in print or they will trail behind the vanguard. We were surprised to learn that what is true for the fields in the headlines is also true for the less publicized ones: nearly everyone needs quick access to recent research, to keep current and to avoid duplication.

Already, in the sciences, several fields access this information quickly. For example, Pl-Net, a service of the American Institute of Physics, has information on papers accepted for publication in a number of physics journals. The Stanford Linear Accelerator has a database with information on preprints, published articles, conferences and symposia in high energy physics. Computer scientists routinely exchange preprints on the Computer Science Network (CSNet).

But even in the sciences, many fields do not have quick access to current research. Astronomers, mathematicians, and physicists in fields outside high-energy physics say they need faster access to preprints. And in the social sciences, this situation is the norm.

What Scholars Tell us About the Scope of the Problem

In virtually every science and social science field, the exchange of papers before publication is a critical part of the research process. At the moment, informal networks are the basis for this exchange. Scholars who are well established with a wide network are inundated with material, some of it relevant, much of it not. An economics professor told me he receives about five preprints every day. This is the basic form of communication in his field; he gets everything he needs, and much more. But younger scholars, or faculty at institutions that are off the beaten track, have to go out prospecting to get papers that have direct bearing on their research.

For example, a psychologist who does interdisciplinary research described to me how she writes to other scholars all over the country to ask for copies of papers they have delivered at a conference, or to get on their mailing list--a time-consuming and undependable way to get information. On the one hand there are the "haves"--established scholars--who are besieged with materials, and requests for materials, and on the other hand, the informational "have nots"--younger scholars, institutions off the beaten track--who spend a lot of time and energy tracking down current research.

Why can't they rely on journals to get this information? First, some of the papers and conference proceedings that circulate through informal networks never get published. The "black market" is the only place they can be found. The pipeline for papers and conference proceedings to be published is very long: in the social sciences, six months to a year for the review process and another year or more for the article to be published. (The average wait for journals published by the American Psychological Association is one to one-and-a-half years.) Therefore a two or three year gap is typical in the social sciences. In the sciences the gap is not quite so big, but it is still significant. As a result, by the time an article sees the light of day, it is no longer "fresh," as one scholar said. Even worse, in a fast moving field like cognitive science, one faculty member told me, "If it's in print, it's out of date."

Working papers and technical reports are equally critical to research. They are generally not published in a formal sense, but issued in a series (in Xerox copies) by academic departments and research institutes. How do scholars get them? It is ironic that, even in economics, they are obtained through what is essentially a barter system: one department agrees to send its papers in exchange for another department's. Often, this informal publication is the only publication they get. Libraries have a hard time collecting these items because they are issued irregularly and not available through the usual channels. Departments sometimes try to collect them, but the information management is often haphazard (although one department has dedicated a closet to its collection). The system of disseminating and organizing this information is almost medieval, except that state-of-the-art copy machines have replaced multitudes of monastic copyists.

Scholars whose work is interdisciplinary are really spread thin, because they are trying to keep up with working papers in several disciplines at once. For example, a sociologist researching student performance in school would at the very least venture into psychology and education; or a geologist studying tidal waves and their effects might need to consult work in engineering and oceanography.

In addition, scholars across the disciplines emphasized how critical it is to be up to date about what other researchers are doing, even before there is a research product, because it allows them to build on other work, and to avoid duplication. Thus it is also important for them to have quick access to information on recently awarded grants. But there is no easy or fast way to get this information, especially in interdisciplinary fields.

The example of the National Science Foundation (NSF) helps to explain why. NSF awards ten thousand grants every year. Getting information on them is possible, but it is not easy, and it is not fast. Biology, for instance, is divided into eighteen different programs; engineering has thirty-one divisions. Just figuring out where to go at NSF, and making sure of contacting all the relevant divisions is very time-consuming.

NSF does make information on its grants available through the National Technical Information Service (NTIS), but NTIS only updates its database twice a year. This is not very helpful if the design of a research project hinges on related work being done by others. Information from other agencies and foundation is even more elusive: there is no central listing whatsoever of grants awarded.

To summarize the scope of the problem with research in progress: the quantity of research today is huge; information about it is hidden; and getting it is horrendously complicated. What can we do to solve the problem of research in progress?

- Work with scholarly associations, publishers, and granting agencies to collect this information and make it available.
- Develop central, comprehensive databases (or one database) with research in progress information on a national network.
- Provide timely access to this information.
- Include abstracts and contact information.

What would be the benefits? Scholars across the disciplines asserted that we would:

- Improve the way research is done in their field by opening it up.
- Streamline the process of getting information on research in progress.
- Reduce duplication.
- Help researchers to design their research projects.
- Provide equal access to all, regardless of career stage or institutional affiliation.

COMPUTER FILES

Our question was, what types of computer files (or in library language, machine-readable data files) do scholars use, and how important are they to their research? As you might expect, they use

all types. A computer file might have a set of numeric values that can be manipulated by a statiscal routine: time series data or longitudinal data. It might have text: a Shakespeare play, or the *New York Times*. It might contain other types of data: analyses of rock samples, or genetic sequences. Whatever they contain, there is no question that, in the sciences and social sciences, computer files are as basic to research as pen and paper.

What Scholars Tell us About the Scope of the Problem

The number of computer files is huge, and growing. Private data sets, developed by individuals or teams or researchers, and not generally available to others, are one important segment of the computer file universe. As the use of computers in more and more sophisticated ways for research grows, so does the number of private data sets.

What kinds of things are we seeing? An anthropologist develops a database of the indigenous languages of Latin America. An astronomer develops a database of pulsars. A group of political scientists assembles data on the political, economic, and social development of contemporary nations. This is happening in all the disciplines.

The number of publicly available data files is even larger. Again, there are many sources. The government is probably the leading producer of statiscal data; the Bureau of the Census and Department of Energy are two of the most prolific. Academic research is another source; an example is the Panel Study of Income Dynamics, which is assembled at the Institute for Social Research at the University of Michigan. Intergovernmental organizations, such as the United Nations and the Organization for Economic Cooperation and Development, also produce data files.

The data files they produce can be enormous, and enormously complex. A good example is the Survey of Income and Program Participation. It contains data from a series of interviews--nine per person--with sixty-four thousand people. The 1984 survey alone takes up twenty-seven tapes. It is easy to see that these projects require a great deal of effort and money. Obviously, to justify an investment of resources of this magnitude, data files must be put to use by many different projects and researchers.

But they are not always used as widely as they could be, because information on them is hidden. For private data sets, the problem is critical. There is virtually no information at all about their existence or location. The professional grapevine is about the only source--but it is not terribly effective. Thus, I found myself telling a scholar about a database project that had direct relevance to his work.

In the case of public data sets, the issue is more complicated. Many government agencies make information on data sets they generate themselves accessible. An economist using the Gross National Product (GNP) or balance of trade data will have no problem locating and

obtaining it. But scholars whose needs are less straightforward report that they have difficulty finding what they want. The earth sciences are a good example. Although there is public access, earth scientists themselves have difficulty knowing what is to be found in an agency like the National Geophysical Data Center. Its data collections cover a vast range--from tide gage records to marine mineral data to solar flare data. But how do scientists find out about them? They call up their contact, if they have one, or get on the agency mailing list. But there is no central source or way to get an overview of what the possibilities are.

Files produced in an academic setting are even more hidden from view. The Inter-university Consortium for Political and Social Research (ICPSR), is a very important resource. It archives and disseminates social science research data, and its records are available on the Research Libraries Information Network (RLIN). But there are over fifty archives in the United States alone that deal with social science data, and most of their holdings are not available on a national bibliographic network. There are even more archives in Europe. Most archives publish newsletters and lists of holdings, but keeping track of developments at all those archives is a job in itself. In the sciences, there is nothing even remotely comparable to ICPSR.

Data files are often hidden for another reason: even when researchers are aware of a file's existence they often do not know that it contains information relevant to their topic. For example, how can the researcher find the 350 studies in the United States that provide estimates of religiosity, particularly when some have titles like "Quality of Employment Survey," and "General Social Survey"?

The scope of the problem with computer files does not end with getting information about them. Researchers also identified a host of horrendously complex questions connected with the use of computer files once they obtain them:

- Documentation is critical: researchers must have it, in order to know what is in the file and how to use it. However, documentation is not always available along with the file, or is often inadequate.

- The collecting, coding, and structuring of data are not standard. Thus, each file is a new Pandora's box, and a great deal of work has to be done before a researcher can even begin to use it.

- Confidentiality is sometimes an issue. Some files are so confidential that only the agency that has assembled them can use them. Consequently, resources of great potential value for research often lie fallow.

What can we do about it?

- Provide better bibliographic information on national networks.

- Develop better facilities for archiving and dissemination of files.

- Provide more detailed information on the contents of data files.

- Encourage standardization in data collection, coding, and file structure.

- Make some files available on national networks. A few of the possibilities are: often used manuals in physics and chemistry; the Index to the Foreign Broadcast Information Service; star catalogs; climate data; seismological data; and bathymetric data.

The benefits are evident. Files would be easier to get, and easier to use. Thus, we would get a far bigger return on the enormous investment made.

It is not enough to say that the problems with research in progress and computer files are huge, hidden, and horrendous. To bring order out of chaos: this needs to be our mission. Otherwise, computer files will just gather dust, and researchers will not get the full benefit of work done by others.

Access can be quicker and more effective than ever. What will make it happen is librarians who are willing to occupy the lead wagon in settling this new frontier. Not only can we meet the challenge of the computer age; this is our eminent domain.

SUMMARY OF THREE NON-BIBLIOGRAPHIC OPERATING ENTITIES

C. James Schmidt[1]

Earlier today Professors Bell and Olson described - from a scholar's point of view - two separate projects. Each project involves non-bibliographic data - in both cases non-textual as well. Both projects involved The Research Libraries Group, Inc. (RLG) as a collaborator and as the network environment in which the data is or will be shared. In the next few minutes I want to discuss three things: (1) describe the larger context of which the two projects you have already heard about are a part; (2) describe the other two projects which were part of the original set of four; and (3) share with the members of the Library of Congress Network Advisory Committee (NAC) several of the questions which have arisen at RLG about the initial set of four projects as well as about future undertakings.

The Program for Research Information Management (PRIMA)

In November 1986, the RLG Board authorized a general program of projects directed at the diverse array of information resources used in research, with special reference to resources not "managed" by existing schemes in libraries and elsewhere. This program was named the Program for Research Information Management (PRIMA). In so doing it was recognized that the kinds of information used in research has grown and will continue to. While scholars continue to depend, in varying degrees, on resources available through and in libraries, there are more and more relevant information resources not only not in or through libraries but also not either bibliographic or textual. Hence, a pilot set of initiatives.

The Four Initial Projects

The initial set of projects chosen ranged from bibliographic

---

[1] C. James Schmidt is vice president of the Research Libraries Group, Inc. (RLG) and director of the Research Libraries Information Network (RLIN) in Mountain View, California. This paper is based on a presentation (by Constance Gould for Mr. Schmidt who could not be present) given on March 30, 1989, at the Library of Congress Network Advisory Committee Meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of RLG or RLIN. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

through textual to numeric and beyond to image data. This initial set was complemented by a needs survey which my colleague Connie Gould described. (Members of NAC received a copy of the Humanities survey after our last meeting.) You have already heard Professor Bell describe the numeric project - the Medieval and Early Modern Databank (MEMBD). A prototype version of the MEMDB which runs on a personal computer (PC) was announced and released last fall. A mainframe version is currently being designed.

RLG is currently actively seeking development partners for the implementation of the completed design of the Geographically Referenced Information (GRIN) project about which Professor Olson spoke.

The other two projects in the initial set were much closer to the bibliographic end of things. First, machine-readable data files (MRDF's). Six university members of RLG (Cornell University, Dartmouth College, Florida State University, New York University, Northwestern University, and Pennsylvania State University) were selected to participate in a project to develop models for selecting, organizing, preserving and servicing MRDF's. Three institutions focused on selection, acquisition and organization (i.e. cataloguing) issues (Dartmouth, Florida, Northwestern); the other three focused on service issues. As a complement to the activities of these six universities, funds were provided to the University of Michigan's Catalog Department to catalog or upgrade records for MRDF's in the collections of the Interuniversity Consortium for Political and Social Research (ICPSR). RLG's MRDF project will conclude this summer with a published report and an invitational conference in September.

The other project in the set of four involved the Modern Language Association and created the Research in Progress Database (RIPDB). I mentioned this project briefly during my remarks at the joint NAC/EDUCOM meeting in December 1988.[2] This database contains citations and abstracts of articles in the areas of language, literature, linguistics and folklore accepted but not yet published by fifty-two participating journals. In addition the database contains records of publication grants awarded by the Division of Research Programs at the U.S. National Endowment for the Humaities (NEH). This year the National Council for Research on Women will begin to contribute records on research in progress by and about women. Discussions are underway to extend the disciplinary scope of this database to include the social sciences and the physical sciences.

RLG has created an advisory committee of scholars which is evaluating this initial set of projects and making recommendations for the future directions of the PRIMA program.

---

[2] The proceedings of the December 1988 NAC/EDUCOM meeting, published as Network Planning Paper No. 18, titled *Connecting the Networks*, include Mr. Schmidt's remarks in a paper "Access to the Research Libraries Information Network Database via the National TELENet." [Ed. note.]

Questions and Issues

It should be no surprise to members of NAC that the RLG experience with its four initial projects has raised a lot of questions. Some of these are: how is a project, i.e. a database, selected? This question is not only one of who decides but also one of what is the library's role in affecting the choice. Some of the answer lies in the governance structure of the organization making the choice in this case RLG. But the larger issue of the library's role and scope as a function and as an institution is at issue here too.

Who pays - for one-time development; for continuing operating expense?

The initial set of projects received their start-up funds from foundations. It is probably not reasonable to expect the philanthropic sector to be a perpetual source of venture capital for such undertakings. Furthermore, most bibliographic networks are unlikely to be able to generate such venture funding from their earned revenues. So where will the start-up funds come from? Presumably, usage charges will recover operating expenses, but then the users who have get and those who haven't don't. A dichotomy between information empowered and information impoverished is not new, but it isn't comforting to see the dichotomy recreated again and again. There are also new cost centers in the equation now - the producers of these new resources, e.g MEMDB.

Conclusion

There are several more issues related to ownership, standards, unmediated (by a library) end-user access. As the survey reported by Ms. Gould suggests, the variety of information resources created by and used by scholars in large and increasing. As the technology of networks we talked about in December gives individuals the means of unmediated access, the members of NAC must address the issues of what can and will be accessed.

BEYOND BIBLIOGRAPHY:
Creating the Rosetta Stone for the Twenty-first Century

Kenneth E. Dowlin[1]

Alvin Toffler contended in his book *The Third Wave*[2] that libraries are second wave institutions. In fact, there is only one reference to libraries in the index of that book. His conclusion convinced me to prove that his assessment was wrong. Libraries can be focal institutions in the Third Wave. We can actually be "New Wave" institutions. The key, as John Naisbitt states in *Megatrends*[3] is networking.

The San Francisco Public Library system (SFPL) provides a good example. In November 1988, the voters overwhelmingly approved the expenditure of $109.5 million in bonds for a new main library and branch library renovation. We intend to design not just a library for the next ten years, but an institution for the next century. This is a major task since we will be redefining the SFPL, and perhaps, public librarianship for major urban public libraries as well.

The redefinition of a major institution is a tremendous task so I have chosen mythology to name the programs within the over-all project. The **Alexandrian Project** is the project to design and construct the building. The **Herculean Project** is the project for cleaning out the Aegean stables. That is to catalog and index for access the estimated 10 million items in SFPL's inventory. The **Mercury Project** is the project to create the electronic network that connects the SFPL internally and externally through electronic technology.

San Francisco's cultural diversity, provides an opportunity and

---

[1] Kenneth E. Dowlin is director of the San Francisco Public Library in San Francisco, California. This paper is based on a presentation given on March 31, 1989, at the Library of Congress Network Advisory Committee meeting in Washington, DC. The opinions expressed herein are the author's and do not necessarily reflect those of the San Francisco Public Library. The author would appreciate the courtesy of notification of any use or reproduction of this paper.

[2] *The Third Wave*, by Alvin Toffler. New York: Morrow, 1980.

[3] *Megatrends*, by John Naisbitt. New York: Warner Books, c1984.

challenge. I intend to redefine the public library in San Francisco. To do so, I have had to establish a vision of the library of the future. The vision is being communicated via a strategic plan and a video tape.

We plan to build a macro information, knowledge, and communication center that will include state, national, and international connections. It would be short sighted, even foolish, to attempt to build a modern center purely on a local basis. In this manner, the public library will be able to stand not only as a monument to learning and creating, but as the key to unlocking knowledge in the cede century.

I contend that the major public library can become the Rosetta Stone of the twenty-first century. Why a public library? The Rosetta Stone was not an academic exercise. It stood at the side of a major crossroads of a major civilization for all to see. A monument to knowledge that everyone could see, and admire--if not understand. It provided the key for later scientists attempting to unlock the past,--and perhaps, today it can serve as a symbol for the future.[4]

I don't have a picture of the Stone of the future. I can only d scribe it by the attributes that must be present. First, it must have wholeness. It has taken me over twenty-five years to discover what is unique about public libraries. It is not that we have books that are available in many places. It is not that we have story hours, many schools have story hours for children. The key is that public libraries have whatever materials are necessary to meet the knowledge, information, and reading needs of the individuals in the community. The whole is dramatically greater than the sum of the parts.

I illustrate by referring to telephone numbers. The SFPL is a major resource for finding telephone numbers around the world. We have not only the microfiche directories for the United States, and phone books from forty-five cities around the world; we have the Minnetell terminal that is connected into an online directory in Paris. It takes all of these to provide a telephone directory system for the user. For wholeness to exist, the system must be multi-lingual, multi-cultural, multi-national, multi-format and neographic. (Neographic is a term that I have used to describe the collections that include not only the traditional materials - graphic - but those that are electronic or machine readable: they are neographic).

The Stone should be magnetic. That is it should attract information and knowledge, facilitate the creation of knowledge from information, and provide archives regardless of format. If truly powerful it would attract ideas and concepts.

---

[4] The Rosetta Stone is a black basalt tablet, found in 1799 at Rosetta in Egypt. It bears parallel inscriptions in hieroglyphic demotic characters and in Greek. It provided the key to the deciphering of ancient Egyptian hieroglyphs. [Ed. note.]

The Stone should have connectivity. Networking is integral to expanding the information and knowledge base. The Stone will be not only an access point, but a node, and a switching center.

The Stone should have directness. It should guide the user to the path that is needed, provide "hypermaps", "hypercatalogs", or intellectual pathways. Intelligent gateways should be profuse and macro authority control should provide intellectual connectivity.

Transparency is important. The methodologies should encompass data, information, and knowledge. Perhaps wisdom. The results of a search, or pathway, should be very specific to the needs of the user, but must be provided in context. In fact the public library is one of the few institutions that can provide the context for civilization-the cultural memory bank. This is extremely important in the rush of Madison avenue to push electronic information systems.

SFPL has a long way to go to develop the Rosetta Stone. A start has been made. We do have funds committed to the development of the electronic systems to enhance the graphic materials. I anticipate that over $8 million will be spent on electronic systems over the next five years. The Library of Tomorrow project supported by Apple Computers Inc., will produce a MacIntosh based knowledge directory that will communicate in Chinese, Japanese, Vietnamese, and English as well as visually and graphically. It is our first venture into the design of the multilingual and multi-cultural online public access catalog. The San Francisco Connection is a project that is serving as the kernel for development of an online electronic mail system, and electronic message system, and an online community resource data base system. In addition, it is providing librarians to supply the information for decision making at San Francisco city hall.

Obviously, there are many networking implications for such a system. If the Rosetta Stone can be re-created at the local level it makes sense to connect such local libraries. It should then be possible for a person to approach the local library for information about a different city and be connected with the appropriate library.

I can illustrate the potential with an actual reference question. Recently, I requested information on events of interest to tourists in Paris during the bicentennial celebration of the French Revolution for a San Francisco magazine. My colleague at the Bibliothèque public d'information in Paris responded the next day with an extremely current list of events planned for the summer related to the bicentennial. I used ALANET, the electronic mail system of the American Library Association, for the communications. I routinely use ALANET to communicate with colleagues in the United States as well as France and United Kingdom.

An online network of libraries around the world would become the Global Village Library and could be a true Third Wave institution. I have been discussing such a network with my peers at the major public libraries in California, and the subject has been brought up at a meeting of the directors of the major public libraries in the United

States. I plan to pursue this type of networking at the conference of the International Federation of Library Associations (IFLA) in Paris during August 1989.

If the United States intends to lead a global economy in the information age (especially competing with countries that focus their resources) the Library of Congress must be the hub of "Knowledge Network USA". (I don't say this lightly, but a colleague mentioned to me the other day that at the Library of Congress the sum of the parts seems to be greater than the whole). The Library must be a focal institution in the global village library. This committee must be the conscience of the Library's involvement--leadership--for networking. We as a profession can not allow the Library of Congress to be on the outside. If anyone has the knowledge to advise the Library, it must surely be this group.

NON-BIBLIOGRAPHIC DATABASES IN THE NETWORK CONTEXT
Meeting Notes and Discussion Comments

Ronald F. Miller[1]

Bell[2]

Professor Bell is concerned with the problems of sharing non-bibliographic data which is produced by individuals or small groups of academic researchers. He is manager of the MEMDB (Medieval and Early Modern Data Base), a project within the Program for Research Information Management (PRIMA) of the Research Libraries Group, Inc. (RLG, see Gould/Schmidt below). It currently consists of raw, sometimes tabulated numerical data, in this case, of currency exchange tables. It can be accessed by Research Libraries Information Network (RLIN) users. An off-line version is available for $250 and is usable on an IBM personal computer (PC).

Bell's goal is to provide access to these data with a simple interface in a format that would allow local data manipulation. His input problem was immense (expense unknown), since such source documents are disorganized and handwritten, and the data are extensive. Rather than creating a huge index to the data, he uses pointers between data elements to permit post-coordinated retrieval. He considers this subsystem a model for other research-oriented databases which probably would not be otherwise commercially available because of the relatively small market and high data collection and entry costs. Because RLIN shows library holdings, the location of related bibliographic references can be found.

A large problem presented to libraries which serve as outlet sites for this kind of information is the lack of interpretive skill needed to instruct potential users about content and organization of the database. Rapidly changing data also frustrate users because paper documentation can't be kept up easily, so other means must be created.

---

[1] Ronald F. Miller is the executive director of the Cooperative Library Agency for Systems and Services (CLASS) in San Jose, California. These notes reflect the program portion of the meeting and synthesize all seven speakers' presentations and related discussions.

[2] Please refer to the meeting agenda (Appendix A) for speakers' full names and affiliations and to the individual papers reflecting most speakers' presentations.

These problems have not yet been solved.

Olson

Professor Olson deals mainly with image analysis and interpretation of photographs recorded from airborne cameras and from geo-sensing satellite telemetry. He is particularly concerned about map collection management and the necessity for preserving historical maps of the same areas for chronological and interdisciplinary interpretation. He related that it was a historian, not a geographer, who pointed out that a discontinuity in a photograph was, in fact, the 150-year-old residual image of the Oregon Trail, not the result of a telemetric or mechanical quirk. Olson also illustrated how land use (and abuse) is documented for legal purposes by comparing a chronological series of aerial photographs of the same land area. "Old maps," he observed, "may be more valuable than new ones."

Olson described several problems: (1) security (some telemetric images are classified by the federal or private enterprise--oil companies, for instance); (2) disorganization (maps are difficult to locate - one has to use the "old boy" network); and (3) multi-data, multi-source correlations create valuable information from such data and so far this is mostly a human function. Such images are very important for ecological monitoring (toxic waste diffusion, for instance). He is also worried about the behavior of researchers, who, spoiled by easy computer access, will not take the time and effort to dig into other sources for information. A member of the audience pointed out that this problem exists in all types of libraries as well and is creating a population of lazy information seekers.

His final question was: "how will networks deal with multi-media output (i.e., a dissertation with two maps appended) with acceptable resolution and cost?" Olson feels that RLIN holds potential as a common interface to such research databases.

McDonald

Mr. McDonald's company is involved mainly with Compact Disk-Read Only Memory (CD-ROM) products, Local Area Networks (LAN's), and database maintenance. He defined "non-bibliographic" as databases which are full text, numeric, or graphic in content. He made the point that numerical databases are not usable unless the data can be analyzed easily by the user. He believes that the work station level of a network is probably the best place for resident analytical software since online communication costs are minimized. He also noted that graphics databases incur high storage and transmission overhead and relatively slow access speed. He sees the advent of more powerful spreadsheet and text manipulation software; hybridized combinations of network and local work station resources will continue to evolve so that the distinction between them will become transparent to the work station user. McDonald also observed that graphics-based work stations, particularly in the library market, will continue to be quite expensive and therefore will probably not be acquired by the average library very soon.

Williams

The Data User Services Division of the Bureau of the Census is a major publisher of machine readable data files (MRDF's) in numeric form: about three thousand titles are published per year. This government agency produces many census-like surveys for the government and disseminates the tabulated data, primarily through the Depository Library system. Anyone can resell their products, with or without adding value, because they are in the public domain. Over five thousand tapes have been sold since the mid-sixties. CD-ROM output has been tested since June 1986 and currently occupies seven compact disks (these may be floppy diskettes) including access software in the dBASE III [database] applications format.

This trend has created problems for public access to these data since many federal depositories have neither work stations nor skilled staff to deal with CD-ROM or computer tape data. Williams feels that it is the libraries' responsibility to improve their staff skills without federal support. These media are distributed, along with the paper reports, through the U.S. Government Printing Office (GPO). Networking aspects were not discussed by Mr. Williams.

Shaw

Mr. Shaw described the capabilities and growth of the multi-service local network offered by the Colorado Alliance of Research Libraries (CARL). Its non-bibliographic files are all textual, including an online journal article location service called "Uncover," various facts files of regional interest, *The Electronic Encyclopedia from Grolier*, selected Wilson indexes, *Denver Business Journal*, and a ride-sharing service. Some of these are not publicly available yet. CARL is a public access system which, through its six installations, serves over seventeen hundred terminals, with over forty thousand uses per day. Currently, including local libraries' Online Public Access Catalogs (OPAC's), twenty-seven files are on the Denver installation and connections are being developed with other systems, such as the University of California and the California State University system. All nodes can access each other. No numeric or graphics files are offered yet. Computer-to-fax capability is being developed to support inter-library loan of journal article abstracts and full texts. The current file size is over three million records, seven hundred and fifty thousand users in forty-four libraries. "Uncover" is expected to have available ten thousand journal articles indexed online within the past year or so. About five thousand of these are available now.

Shaw pointed out that database licensing, copyright control, and database selection criteria are key issues for CARL now. So far, the alliance hasn't had to provide much human value-added support to users, thereby keeping costs (and changes) down.

Gould[3]

Ms. Gould reported on an RLG survey of information use behavior by scholars, which resulted in the building of the Research in Progress (RIP) file on RLIN. She noted that the quantity of reportable research is high, but the information about it is hidden and sparse, and the access problem is great (much of the research data isn't captured in electronic form yet). The Inter-university Consortium for Political and Social Research (ICPSR) database forms the core of the RIP file project.

The four projects which RLG has undertaken in its Program for Research Information Management (PRIMA) include Professor Bell's MEMDB (see above), Geodata (see Olson above), data files from six universities, and the RIP file cosponsored with the Modern Language Association. The RIP contains interim research reports about language, linguistics, folklore, women's studies, and some National Endowment for the Humanities (NEH) research.

Dowlin

Kenneth Dowlin, the director of the San Francisco Public Library (SFPL), presented a vision of SFPL services and their possible impact on the Bay Area in the twenty-first century. Having floated a $109 million bond issue for a new library facility, Dowlin sees SFPL as a leader of "new wave" libraries. He is particularly focusing on improving access to information and knowledge in a multicultural, multilingual context. Several projects are under way with Apple Computer and Pacific Bell, one of which is a high technology video about the future library, which he played for the Network Advisory Committee. His vision portrayed the library as the information and knowledge center of the metropolis--a manifestation of the "global village" library envisioned by Marshall MacLuhan a few decades ago, providing electronic and communication services of many types, in vernacular languages. San Francisco is a city in which there is no longer a majority of any race, ethnicity, or culture. In a memorable phrase, Mr. Dowlin said: "we are drowning in information but starving for knowledge." He intends to hold a meeting about networking the major public libraries in California to try to expand his vision state wide.

Dowlin concluded with the observation that the Library of Congress should be the focal institution in the global village network and that the Network Advisory Committee must be its conscience.

Discussions

"What should we [librarians, information providers] do about the future lack of control of databases," asked Henriette Avram. "How

---

[3] In addition to her own presentation, Ms. Gould substituted for C. James Schmidt and read his paper following her own presentation. Mr. Schmidt's paper is included in the proceedings. [Ed. note.]

do we deal with getting to the future?" These words expressed everyone's feelings and a number of comments were given and questions raised during the ensuing discussions.

- Personal research must be captured and made accessible.

- A common user interface which can be used by scholars and the literate lay public to access and display bibliographic, full text, numeric, and graphic data files is needed.

- Librarians must learn how to cope with the multitude of forms in which information is available.

- Some new media, such as CD-ROM, appear to reduce user access by virtue of their new technology and the policies which surround them. (Note: a recent survey says that there are 171,000 CD-ROM drives in the world today--hardly enough to be commonplace yet.)

- Issues of standardization and information selection remain unresolved.

- Scholars complain about slow dissemination of research results through "normal" channels. They have status and promotion problems if they publish results in non-standard forms.

- There is a cluster of problems imposed on libraries and their network organization by the recent availability of "low cost" local stand-alone systems and their relationship to bibliographic utilities. Resource sharing seems to be a casualty and network revenue is reduced by migration of network users to some of the seemingly lower cost choices.

- DIALOG has over twenty million non-bibliographic records in its system and demand is growing.

- A new generic label for "non-bibliographic" databases that removes its library-centric negative connotation should be considered.[4]

A number of recommendations for action emerged from the group discussion. Some reflect the initial comments and questions raised by NAC members, others introduce new concerns. Most suggestions can be grouped into "more information," "stave off chaos," and "contribution." Some of them are:

- develop a format for database collections;

---

[4] Bill DeJohn, director of the Minnesota Interlibrary Telecommunications Exchange (MINITEX), and Anita Anker Branin, assistant director for document delivery at MINITEX and DeJohn's alternate to NAC, were asked to look for a more positive term for databases that are non-bibliographic.

- develop generic standard description (standards cannot be developed in a vacuum);
- develop a generic standard reporting format (there exists a very diverse universe of information and some of it is not relevant to others);
- promote availability of such a format;
- encourage and make use of networking protocols which are already reaching beyond bibliographic databases;
- change NAC's mind-set from bibliographic networks to information networks;
- try <u>not</u> to create a central directory;
- use electronic bulletin boards;
- create a "mega base" for neographic [see Dowlin, "Beyond Bibliography"] records, i.e., a critical mass of records;
- study technical and political feasibility of databases;
- use Federal resources to develop such a feasibility study;
- stay informed about text encoding activities under way in the humanities;
- gather information about databases and how they are disseminated;
- seek to enlighten NAC members about where non-print information is and how it is recorded;
- continue the linking of super computers (already thirteen of them link universities, according to EDUCOM); and
- establish a subcommittee to pull together into a single statement a description of the situation as it seems to exist, where all of this is leading (more information is needed to develop strategies about how to deal with the situation), and finally, define a hoped-for goal.

SUMMARY OF BUSINESS SESSION

The agenda of the business session at the Library of Congress Network Advisory Committee (NAC) meeting was brief. Several items were on the agenda: new representatives, NAC meeting topics, a letter to Congressman Kastenmeier, the Membership Subcommittee report, planning for a pre-White House Conference, and the next meeting's date and place.

New Representatives

Henriette D. Avram, chairman of NAC, opened the March meeting by welcoming all attendees. She extended a special welcome to a new NAC member organization--the Capital Consortium Network (CAPCON) with its executive director, Dennis Reynolds, as its representative. Mrs. Avram reported a number of changes in the NAC membership representation that had occurred during the last few months. Lou Wetherbee, who represented the AMIGOS Bibliographic Council, Inc., left AMIGOS and Bonnie Juergens, the new executive director at AMIGOS will be the new representative. Bill Studer, who represented the Association of Research Libraries (ARL), was replaced by Duane Webster, the new executive director at ARL. Because Mr. Webster could not attend he sent his alternate, Prudence S. Adler, instead. James Riley who represented the Federal Library and Information Center Committee (FLICC) since 1976 on NAC, retired recently from the Library of Congress; his alternate, D. Lee Power attended, no new representative was named so far.

NAC Meeting Topics

1. Networking Activities in the States

Correspondence between Mary Ellen Jacob and Clarence Walters (both at OCLC, Inc.) concerning networking activities in the states was distributed prior to this meeting. The main document was the revised draft version of "State and Commercial Bibliographic Activities and their Effect on the Bibliographic Utilities." Several NAC members agreed with Walters' conclusion that this topic warrants another set of discussions by NAC. It was felt that another paper should be prepared and distributed in advance of the NAC meeting - possibly a survey of the 1989 state of affairs regarding state databases. [The meeting topic, decided upon in the following months, became more focussed and will concentrate on the impact of the local library systems on state and national networks. The date for the meeting was set for March 28-30, 1990. Ed. note].

2. Review Process of Adopting the Vision Statement

Another meeting topic suggested is on the review of the process of adopting the Vision Statement. NAC has knowledge of twenty, perhaps there are more. It should be discussed why only twenty organizations have adopted the statement. Such a meeting would also allow NAC members to review the list of action items still to be

completed in accordance with the agreement at the conclusion of the December 1986 program session. [For a list of action items see *Network Planning Paper* no. 15, "Nationwide Networking: Proceedings of the Library of Congress Network Advisory Committee Meetings July and December 1986," p.22-26. Ed. note.]

Letter to Robert W. Kastenmeier

A letter, signed by the Librarian of Congress, Dr. James H. Billington, was sent to the Honorable Robert W. Kastenmeier, chairman of the subcommittee on Courts, Intellectual Property and the Administration of Justice, House Judiciary Committee, at the U.S. House of Representatives. The letter informs Mr. Kastenmeier and his subcommittee of the two program sessions, held by NAC, on the topic of intellectual property rights in an electronic age and related issues. (See Appendix B.)

Membership Subcommittee Report

Lois Ann Colaianni, associate director of library operations at the National Library of Medicine and chair of the Membership Subcommittee, reported on a new application for membership in NAC: EDUCOM's Networking and Telecommunications Task Force (NTTF). The NAC membership approved EDUCOM's application by voice vote following the subcommittee's recommendation. Ronald Larsen, the associate director for information technology at the University of Maryland's McKeldin Library, will represent the organization and begin attending NAC meetings in November 1989.

Pre-White House Conference Planning

Mrs. Avram reported on planning activities of NAC members for a second "Networks for Networkers" conference to be held sometime in 1990 in anticipation of the second White House Conference on Library and Information Services. Barbara E. Markuson, the executive director of the Indiana Cooperative Library Services Authority, chair of the 1979 pre-White House Conference and editor of *Networks for Networkers*, was asked to join the planning group. The first meeting date will be in April 1989.

Next Meeting and Adjournment

The date for the next meeting was set for November 29-December 1, 1989. The topic of the meeting will be document delivery and how the newer technology, such as facsimile transmission, has influenced and is changing current practices in document delivery. Mrs. Avram appointed Charles P. Bourne, the director of the General Information Division at DIALOG Information Services, Inc to chair the program planning. Other members on the planning subcommittee are Lois Ann Colaianni, NLM; Dennis Smith, director for library affairs at the University of California, Berkeley and Henriette D. Avram.

Mrs. Avram adjourned the meeting at noon on March 31, 1989, after thanking all for their active participation. She expressed special thanks to the program subcommittee chair, Sandy Paul, and the other members of that committee for the selection of stimulating speakers addressing the complex and issues of "non-bibliographic" databases in the library networking environment.

LIBRARY OF CONGRESS NETWORK ADVISORY COMMITTEE

MEETING AGENDA
March 29-31, 1989

Wednesday, March 29 8:30 pm	**BUSINESS SESSION** Presideing: Henriette D. Avram, Library of Congress o Report from the chair o State and commercial bibliographic activities... o Membership Subcommittee report
Thursday, March 30 9:00 am	**PROGRAM SESSION** Henriette D. Avram, chairman Welcome and introduction to goals of meeting Sandra K. Paul, S.K. Paul Associates Chairman, program committee Introduction of program Rudolph L. Bell, Rutgers University User perspectives and requirements: creator of non-bibliographic databases has to share data with others Charles Olson, Jr., University of Michigan Roles of societies and associations: e.g., a geographic database Dennis McDonald, Online Computer Systems, Inc. Commercial sector: non-bibliographic databases via networks Forrest Williams, U.S. Bureau of the Census Government sector: Census Bureau database via networks Ward Shaw, Colorado Alliance for Research Libraries Regional level activities: sharing of non-bibliographic databases Constance Gould, The Research Libraries, Inc. National level: RLG's assessment of non-bibliographic needs C. James Schmidt, The Research Libraries, Inc. Summary of three non-bibliographic operating entities (databases)
Friday, March 31, 9:00 am	Kenneth E. Dowlin, San Francisco Public Library Futuristic look at dissemination of non-bibliographic databases via networks Ronald F. Miller, CLASS, moderator Final plenary session
11:00-12:00 noon	**BUSINESS SESSION** (Cont.)

March 1 5 1989

Dear Mr. Kastenmeier:

The Library of Congress Network Advisory Committee devoted two program sessions (April 1987 and March 1988) to the topic of intellectual property rights in an electronic age and related issues in the library network context. The Committee, a body representing some thirty organizations of the U.S. library and information community, both profit and non-profit, has been meeting since 1976 and addresses current issues of concern to the library and information community.

At the April 1987 meeting Michael kemington of your staff was invited to speak. At that time, he indicated that your subcommittee would be interested in the results of the Network Advisory Committee's deliberations on the subject of intellectual property. He felt these observations would provide insight into the copyright problems that arise in the new electronic environment of the library and information community. One thing is clear; new technology has a significant impact on the informational and copying activities of libraries. Therefore, a strong case can be made for expanding the Copyright Office study mandated by section 108 (i) of the copyright law to include these new technologies.

I am enclosing for your Subcommittee's information a background paper prepared for the March 1988 meeting. It points out the critical issues for intellectual property created by the development of new technologies. The second enclosure is a summary of the meeting's discussions and identifies other issues concerning intellectual property rights as well as documenting four major points on which the Network Advisory Committee members reached consensus.

I hope you and your Subcommittee find this information useful. The Network Advisory Committee will hold another meeting on this topic later this year or in early 1990.

Sincerely,

James H. Billington
The Librarian of Congress

Enclosure

The Honorable
Robert W. Kastenmeier
Chairman
Subcommittee on Courts, Intellectual Property
and the Administration of Justice
House Judiciary Committee
U.S. House of Representatives
Washington, D.C. 20515

HDA:sgh
3.6.89

No. 5 Long, Philip L. *Study of Message Text Formats: Bibliographic Search Queries.* (Washington: Library of Congress, 1979). 28 p. $4.00.

No. 4 Network Technical Architecture Group. *Message Delivery System for the National Library and Information Service Network: General Requirements.* (Washington: Library of Congress, 1978). 35 p. $4.00.

No. 3 Buchinski, Edwin I. *Initial Considerations for a Nationwide Data Base.* (Washington: Library of Congress, 1978). 56 p. $4.00.

No. 2 Dataflow Systems. *A Glossary for Library Networking.* (Washington: Library of Congress, 1978). 34 p. $4.00.

No. 1 Butler, Brett. *A Nationwide Location Data Base and Service.* (Washington: Library of Congress, 1978). 66 p. $4.00.

Single copies are available for sale from the Customer Services Section, Cataloging Distribution Service, Library of Congress, Washington: DC. 20541.