Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

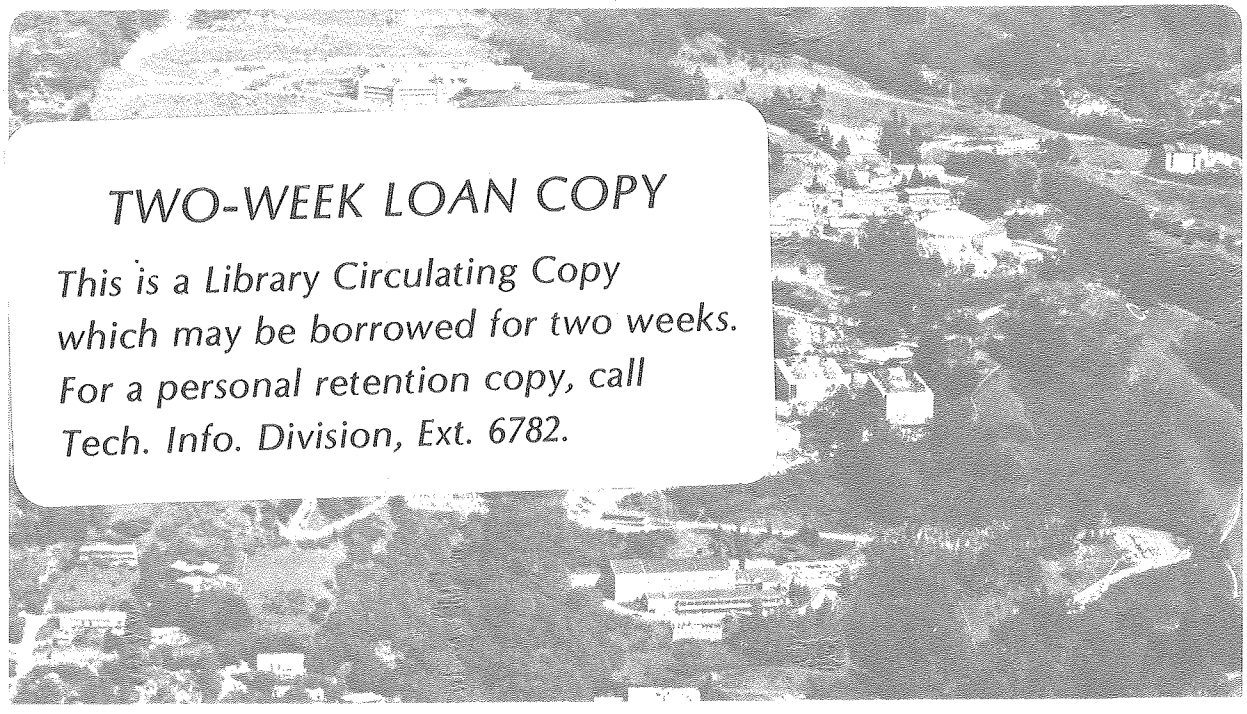
Engineering & Technical
Services Division

Presented at the Fourth Institute of Electrical and Electronic Engineers Symposium on Mass Storage Systems, Denver, CO, April 15-17, 1980

MASS STORAGE SYSTEMS AND LARGE RESEARCH LIBRARIES

James A. Baker

April 1980



Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

DISCLAIMER

LBL-10792

To be presented at the Fourth
Institute of Electrical & Electronic
Engineers Symposium on Mass Storage
Systems, Denver, Colorado,
April 15-17, 1980

MASS STORAGE SYSTEMS AND LARGE RESEARCH LIBRARIES*

James A. Baker

Lawrence Berkeley Laboratory
Berkeley, California

April 1980

* This work was supported by the U.S. Department of Energy under contract No. W-7405-ENG-48.

MASS STORAGE SYSTEMS AND LARGE RESEARCH LIBRARIES*

James A. Baker
Computer Center
Lawrence Berkeley Laboratory
Berkeley, California

April 1980

ABSTRACT

A scheme to reorganize large research libraries is presented. The scheme involves storing infrequently circulated books on a mass storage device in digital form. The proposed system would solve many problems associated with the proliferation of books and serials in research libraries.

INTRODUCTION

In 1962 Kemeny[1] proposed a radical solution to the problem of the proliferation of books and journals in large research libraries. His suggestion involved the funding and construction of a National Research Library which was to be used by scholars and students throughout the country. In the National Research Library books were to be stored in image form on something resembling video tape. Users at remote sites would access books in the central library via communication lines. Kemeny envisaged that the technology to construct such a library would be available by the year 2000. He proposed that planning for the construction of that facility should begin in 1980.

---

* This work was supported by the U. S. Department of Energy under contract No. W-7405-ENG-48.

In this paper we propose a system which is quite different from the one suggested by Kemeny--though it solves the same problem. We assert that this system could be in operation by 1985 and that the technology for its implementation exists today.

Moreover, because of the technology employed, the system that we propose is much less expensive (even in 1980 dollars) than the one proposed by Kemeny. This fact makes the system interesting not only to universities in the United States but also to developing countries who wish to establish new research libraries.

THE PROBLEM

The production of scholarly books and papers is proceeding at an ever increasing pace. The acquisition, storage, and retrieval of these materials requires an increasing portion of the resources of institutions which support large research libraries. For example, the cost of acquiring and cataloging a book at the University of California, Berkeley exceeds $15; this excludes the purchase price of the book. The storage of books requires an increasing amount of expensive floor space. Large numbers of employees are needed to catalog, retrieve, and reshelve books.

Existing research libraries are not particularly convenient for their users. The user is invariably required to go to the library to examine books. In certain instances he may, however, examine the card catalog from a console in his office.

When the user reaches the library he may find that one or more of the items he is interested in are already checked out to someone else. Items in the library may be mis-shelved or be sitting in a graduate student's carrel.

CHARACTERISTICS OF LARGE RESEARCH LIBRARIES

Large research libraries are characterized by:

1. Containing a large number of books.
2. Having only a small proportion of books which circulate on a regular basis.

At Case Western Reserve University, in the fall semester of academic year 1971/72, 19,182 books (not necessarily all different) were checked out from a collection of 1.2 million items[2]. In the academic year 1978/79 at UC Berkeley, 376,152 books were checked out of the central collection of 2,540,271 volumes[3]. Note that this does not include circulation in the undergraduate library or departmental branch libraries. It is interesting to note that, during the same period, 480,504 books were checked out of the Berkeley undergraduate library.

A PROPOSED LIBRARY SYSTEM

We propose a library system in which books which have a high circulation rate (greater than two or three checkouts a year, say) will be retained as books. All other materials would be converted to digital (as contrasted with image) form. In general, departmental branch libraries and undergraduate libraries would be left intact. We propose to dispose of older low circulation items which have been transcribed into digital form. Original acquisitions of new, low circulation items would be in machine-readable form.

Ideally we shall transcribe all library materials (except heavily graphical items), even those high circulation items which are retained or acquired as books.

Books containing a high percentage of graphical material will be excluded. Certain graphical material, such as photographs, will be stored in image form.

The alternative of storing books in image form (as proposed in the Kemeny plan) is rejected for several reasons. The density of material stored in image form is at least an order of magnitude less than the density of material stored in digital form. The retrieval of information stored in image form, particularly the retrieval of subsets, is very much more difficult. Moreover, the quality of reproduction of material stored in image form tends to be inferior to that of material stored in digital form.

The Transcription Problem

There would be no difficulty in the digital transcription of current publications. Such material is generally available in machine-readable form since it has been composed with the help of a computer typesetting device.

The digital transcription of older material presents some problems. We are faced with two alternatives: we must either develop a rather universal optical character recognition device or help to solve the national unemployment problem by hiring a very large number of keypunch operators.

It might be well, while we are solving the transcription problem, to start out by transcribing only new materials. This would solve a substantial part of the difficulties faced by existing large research libraries.

Hardware Elements

Books will be stored on optical disks. A disk containing 10^{10} bits per side could store 2,000 books. These disks would be housed in a device resembling a juke box. A library of two million books would require a juke box containing 1,000 disks. Two or three optical disk readers would be attached to each juke box. Items retrieved from the optical disk would be transferred to a magnetic disk cache system with a capacity of about 2,000 megabytes. Notice that this cache could hold 2,000 books of average size.

In addition to library items currently being examined by users at consoles, the magnetic disk cache system would also always contain a copy of the complete library card catalog. Note that it would be quite feasible to include on each disk all of the library cards for the (roughly 2,000) books on that disk. These cards could then be collated into the main catalog on magnetic disk and onto backup magnetic tapes.

There would be numerous asynchronous communication ports attached to the computer which controls the library system. It is estimated that a machine of the size of a DEC VAX/11-780 or an IBM 4341 would be adequate to operate the system.

The final hardware element in the system would be a xerographic printer with capabilities similar to those of the Xerox 9700.

The Operation of the System

The user of the proposed system would sit at an ordinary computer terminal either in the library or in his office. He could then thumb through the library card catalog. He could request the retrieval and transfer to the

cache disk of any items of interest. From his terminal he could examine these items: he could look at the table of contents, the index, the chapter headings, or various pieces of text. In the event that the item was a short paper, he could read the whole thing while seated at his terminal.

If, however, he desired to have the item, or some subset of it, as a book, to take home and read at his leisure, he could give the computer an appropriate signal and the computer would print a copy of the item for him or, if the item were a high circulation one, cause an existing hard copy of it to be paged and sent to the circulation desk.

When the user went to the circulation desk to pick up his book, he would receive either the specially printed copy or an ordinary circulation type book. In the latter case, he would be asked to return it. In the former case, he would be asked not to return it.

THE COPYRIGHT PROBLEM

Many problems connected with the digital transcription and copying of materials protected by copyrights remain to be solved[4,5]. It is anticipated that when a disk containing copyright-protected works is purchased, that the price will include appropriate remuneration for the holders of copyrights.

It is not expected that the copyright holders will receive compensation when their protected works are read at consoles by library customers. This type of use is regarded as equivalent to checking a book out of the library.

However, when a hard copy of a protected work is printed at a user's request, then it is expected that the copyright holder will be compensated.

IMPLEMENTATION OF THE SYSTEM

Very substantial savings in space and personnel costs will be realized by large research libraries which adopt the proposed system[6]. Already established libraries will realize the greatest savings from obtaining current monographs and journals in machine-readable form. Institutions which wish to establish new research libraries--especially in developing countries--will realize the greatest savings in acquisition costs of old material.

Development Costs

Development costs for the library system, i.e. the part of the system to be installed in the library, are estimated to be $1.2 million. This includes the purchase price of a prototype system including computer, magnetic disk system, one optical disk reader, and a juke box, but excludes the xerographic printer. These items, with the exception of the optical disk reader and juke box, are standard off-the-shelf devices. The digital optical disk and juke box exist as prototypes but not as announced products.

The estimate includes five man years of programming effort, one man year of engineering coordination, two man years of technician time, and one man year of technical supervision. An elapsed time of two years will be required for this development.

Development of an optical disk writing facility--that is one which copies books which are already in machine-readable form to optical disks--is estimated to cost $600,000. This includes the procurement of prototype hardware and the development of software. Elapsed time of 18 months will be required for this phase of the project.

We do not attempt at this time to estimate the effort required to develop an appropriate universal optical character recognition device.

Timetable

Total elapsed time required for the installation of a working prototype in a research library and the creation of a central copying facility is estimated to be five years. This includes the time required for the development described in the preceding paragraphs; in addition, it includes time for legal research on copyright problems, time for negotiation with publishers, and time for training of staff in the prototype installation

REFERENCES

1. J. G. Kemeny, "A Library for 2000 A.D.," Management and the Computer of the Future, M. Greenberger (Ed.), MIT Press, Cambridge, Mass., 1962.
2. F. E. G. Waber and T. Savacevic, A Study of Library Users and Use in a University: Selected Results, Proceedings of the ASIS Annual Meeting, Vol. 11, 1974.
3. V. Pratt, School of Library and Information Studies, University of California, Berkeley. Private communication.
4. H. S. White, Indiana University, Results of a Survey of Publishers' Practices and Attitudes Toward the Authorization and Licensing of Library Copying Under the Terms of the New Copyright Law, Proceedings of the ASIS Annual Meeting, Vol. 15, 1978.
5. National Commission on New Technological Uses of Copyrighted Works, NTIS Report PB271003.
6. H. Fussler and J. Simon, Patterns in the Use of Books in Large Research Libraries, The University of Chicago Library, 1961.