

LIBRARY 2000

Annual Progress Report, July 1, 1992—June 30, 1993

Academic Staff

Jerome H. Saltzer

Research Staff

Mitchell N. Charity

Undergraduate Students

Jeremy A. Hylton
Robert C. Miller
Manish D. Muzumdar
Chad Phillip Brown

Reading Room Staff

Newton M. Loui
Paula Mickevich
Maria T. Sensale

M. I. T. Library Staff

T. Gregory Anderson
Keith Glavash
Lindsay J. Eisan, Jr.

Support Staff

Lisa M. Eastman

1 Introduction

Library 2000 is a research project that is exploring the system implications of large on-line storage. The method of the project is pragmatic, to develop a prototype of an on-line electronic library using the technology and system configurations expected to be economically feasible in the year 2000. Support for Library 2000 has come from grants from the Digital Equipment Corporation; the IBM Corporation; ARPA, via the Corporation for National Research Initiatives; and uncommitted funds of the Laboratory for Computer Science.

The basic hypothesis of the project is that the technology of on-line storage, display, and communications will, by the year 2000, make it economically possible to place the entire contents of a library on-line and accessible from computer workstations located anywhere. The goal of Library 2000 is to understand and try out the system engineering required to fabricate such a future library system. The project's vision is that one will be able to browse any book, journal, paper, thesis, or report using a standard office desktop computer, and follow citations by pointing—the thing selected should pop up immediately in an adjacent window.

Our research problem is not to invent or develop any of these evolving technologies, but rather to work out the system engineering required to harness them in a usable form. The engineering and deployment of large-scale systems is always accompanied by traps and surprises beyond those apparent in the component technologies. Finding things, linking things together (especially across the network), keeping the whole system reliable, and making the system last for a period of decades will all require new ideas.

This is the second reporting year of the project. Last year's report described the overall architecture of the testbed system. This year we report on progress in designing and implementing that architecture, and some closely related issues that we have explored.

2 Overview

The current activities of the project are centered on the building and priming of a testbed—an operational library system filled with live data and used daily by other laboratory members. This testbed is an example implementation of an evolving architecture that exploits technology that we expect soon to be widely and economically available.

Data for the testbed currently comes, or is planned to come, from three sources, each intended to stress the system design in a different way:

- The catalog of the LCS/AI Reading Room. Although small (20,000 records), maintenance and delivery of this collection is an on-going production activity. It thus allows us to insure that we have covered the operational aspects of running a real library.
- The catalog of the M. I. T. Library. This collection of about 600,000 card catalog records is large enough to require careful consideration of how the indexing system design scales. It also allows exploration of how large random access memory might be exploited.
- Page images and text of computer science technical reports. This (prospective) collection of information will provide a test of how the storage service and its replication, updating, and notification strategies scale up in size. As described in more detail in a later paragraph, it will also be the first collection that spans administratively independent libraries, thus providing a test of linking, setting of standards, and more generally of what kinds of cooperation will be necessary or appropriate in future, on-line libraries.

Much of the design work this year was accomplished in a weekly series of group technical meetings that included several graduate students who, though not regular group members, are interested in the research issues raised by this application. These graduate students included Alan Bawden, Thomas Y. Lee, G. Winfield Treese, and Quinton Y. Zondervan. Karen R. Sollins also participated in some of those meetings.

An important aspect of the project, suggested above by the word cooperation, is that Library 2000 is involved in two major cooperative ventures. The first involves working closely with the M. I. T. Distributed Library Initiative (DLI), itself a cooperative venture between two M. I. T. entities, the Library System and Information Services. This group has generated joint proposals, held weekly discussions of items of mutual interest, and discovered opportunities to save work by sharing catalog data and page images. As part of this interaction, Mitchell Charity gave a talk about and demonstration of Library 2000 to a DLI workshop on Radically Electronic Library Services.

The second cooperative venture, called the "CS-TR project", is an inter-university project coordinated by the Corporation for National Research Initiatives (CNRI). The project includes Cornell University, Carnegie-Mellon University, the University of California at Berkeley, and Stanford University. Under this project, each of the universities will obtain, and place on-line, page images of its own computer science technical reports, and make them available to the other participants, thus producing a cooperative library. ARPA funding for this project was arranged during the year and detailed work began in the Spring. The page image collection mentioned above will be the M. I.T. contribution to this cooperative project.

3 Testbed Implementation

The testbed was initially implemented in 1991-92 by Mitchell Charity who, working together with several undergraduate students, is actively extending it in several directions. It currently consists of an image storage service, an indexing service, a telnet client, an X Window System client, and a prototype of the protocol that connects those components. The testbed delivers the three collections mentioned in the previous section (for the moment, the TR collection is represented by a set of searchable abstracts of LCS and AI technical reports.)

As mentioned, during the year, the testbed was expanded and extended. Rob Miller and Manish Muzumdar, undergraduate research students, designed, developed, and implemented the X Window System client for the testbed. Muzumdar also designed and implemented an initial image display browsing interface for the window client. These were then experimentally connected with an server containing a small number of images derived from hand scanning, and from page images taken from Berkeley and the TULIP project (a DLI endeavor).

In addition to these clients, Charity built a testbed programming environment and some gateways. The programming environment uses the Scheme language to allow quick experimentation with new ideas. Two gateways to other information delivery systems have been tested - a TechInfo catalog searching gateway, and a World-Wide Web image server experiment.

Miller added multiplexing management that eliminates memory-space bottlenecks on the number of simultaneous users of the testbed. Muzumdar developed a C library to simplify use of the Library 2000 protocol. Jeremy Hylton, another undergraduate research student, explored several issues related to image handling: compression technology, laser printing technology, and scanning system design. The current status of the scanning system design analysis is described in some depth in a later section of this report. Hylton also developed a gateway that would allow an index to network news to appear as a testbed collection; that work is on hold pending delivery of RAM storage large enough to hold such an index.

Charity, using tapes provided by Thomas J. Owens of the M. I. T. Library Systems Office, installed and placed in service on the Library 2000 testbed system the entire M.I.T. library catalog (about 600,000 MARC records) using a new, large-database indexing engine that Charity and Rob Miller designed and implemented. With modern high-density tape recording technology, this entire database can now reside on a single tape cartridge, and the database in that form has been returned to William Cattey of M. I. T. Information Services to prime the Willow/BRS experimental catalog system, a DLI project.

4 Intellectual Property Issues

Negotiation of the research agreement for the CS-TR project with the Corporation for National Research Initiatives has been a surprisingly significant activity. The reason is that that negotiation has revealed a number of intellectual property and legal questions surrounding on-line technical report distribution that are suitable subjects for research in themselves:

- Where does liability for content problems in a cooperative information system such as the CS-TR project lie, and are indemnification clauses needed in participation contracts? Who is responsible when materials contain embedded, unpermitted copyrights, plagiarization, scientifically invalid data, or dangerously incorrect material? Is it practical or appropriate for a university to vet its materials before distribution? In practice, how much of this class of liability problem is real, and how much is imagined or fantasy?
- What permissions and alerts are appropriate for a university to require of or notification to its authors as part of publishing a technical report?
- To what extent are notices of rights, permissions and warranties necessary or appropriate as part of distribution of university- developed scientific materials?
- Does it make technical or policy sense to attempt to try to restrict use of materials distributed over the Internet and made available to all members of a university? If so, what restrictions are reasonable and enforceable?

In a series of meetings with M. I. T. intellectual property specialists, we worked out a plan for permissions and alerts to go with acceptance of future technical reports, as well as a plan to deal with copyright concerns surrounding existing technical reports.

5 Storage Server and Replication Design

In the Library 2000 model, a major component of the future library will be an independent storage service that may be administratively independent of any discovery (indexing and search) service. The storage service is dedicated to holding and delivering documents; a document consists of page images, as well as a bibliographic description, ASCII text, and maps that relate the ASCII text to the page images. This year we made some progress in defining the desired properties of this service. Two primary design areas were explored: the interface semantics provided by the storage service, and the internal architecture, in which it is assumed that replication with wide geographical diversity will be needed for long-term reliability. Mitchell Charity also developed a prototype, experimental implementation of the transaction, update, and what's-new semantics described below, and tried them out on a

copy of the reading room collection.

Here are some preliminary ideas about this design. First are several criteria that the storage service design should meet. Second are a number of assumptions about the way that the service will work.

The design criteria:

1. The storage service should be able to work on top of any standard operating system, file system, and disk storage system without requiring change or special features.
2. In the library application, instant availability of updates to reading clients is not a priority requirement. Specifically, if the updating mechanism is temporarily unavailable, prospective updates can be queued by the updating client (that is, the curator of the catalog), so long as the time of queuing is bounded.
3. Similarly, it is not essential that all replicas return identical answers with respect to recently changed data. Specifically, it is acceptable that updates are visible at one replica before they are visible at another, so long as the time of disparity is bounded.
4. The intended scale and consequent configuration look like this:

Design center:	1000 maximum-size disk drives.	
	1993: 1000 X 2 Gbytes.	20 Million page images (50,000 books)
	2000: 1000 X 100 Gbytes.	1 Billion page images (2.5 Million books)

Design limit: At least 10 times larger.

At the design center, a 1993 server site configuration might consist of about 20 server processors, each with 50 5-inch Winchester disks and 128 Mbytes of memory, interconnected with an Ethernet. A year 2000 server site would probably have the same number of (faster) processors and (bigger) disks, with more memory and a higher-performance network. The currently planned L2000 prototype is about one tenth the size of the design center. Both in 1993 and in 2000 a design center server site consists of about \$2.5M worth of equipment. In all cases, disk and channel transfer rates are not expected to be a bottleneck; although a library contains a large amount of information, the average utilization rate is low when compared with traditional disk applications such as corporate databases.

5. Simplicity: Because it is difficult to be confident that a large-scale system is actually operating the way one intended, the replication scheme should comprise a small number of relatively orthogonal, independent mechanisms that obviously do the right thing and that can easily be verified to be working when in the field. Thus simplicity

is strongly preferred over a more efficient but more complex mechanism, or even a very elegant, but potentially fragile, design.

6. All interactions among replicas should be accomplished via the standard storage service protocol, rather than by introducing a new protocol or introducing dependence on some other network protocol.
7. The data is assumed to have a lifetime far exceeding the expected lifetime of any particular storage medium and also far exceeding the mean time between major disasters (hurricane, earthquake, civil unrest) that may completely destroy a storage site.
8. All expected events are to be tolerated without loss of data and also logged for analysis of their occurrence rate and pattern. To the extent physically possible, all intolerable events should be detected and reported.

Here are the assumptions about the way that a storage service would work:

- A. The storage service has three kinds of clients: ordinary users, catalog managers, and discovery services. An ordinary user, in possession of a document identifier, presents that document identifier to the storage service, and expects to receive selected parts of the document in return. This user makes use of the read interface of the storage service. A cataloguer presents the storage service with new documents to be stored. The cataloguer makes use of the update interface of the storage service. A discovery service is an intermediary service that provides indexing and searching to allow its own clients to find things of interest. The discovery service makes use of an unusual interface to the storage service known as the "what's-new" interface.

The overall model is that the discovery service uses the "what's-new" interface to keep its document index up-to-date. An ordinary user first presents a query to the discovery service. The discovery service returns a list of documents that may be relevant to the query, and if the user thinks any look interesting, it then asks the storage service for bibliographic information, abstract text, or page images of those documents. One discovery service may provide indexing and searching services for many different storage services, and a single storage service may be indexed by several different, competing discovery intermediaries.

- B. Each client's identity is accurately known to the service, for example by virtue of the client presenting a Kerberos ticket. The semantics of that presentation are not yet defined. There are separate access control lists for reading, adding new records and modifying existing records, and use of the what's-new interface. The semantics of management of these lists is not yet defined. There is no firewall between transactions; anyone with permission to put and modify can join in, commit, or abort any outstanding transaction.

- C. The data stored is one uniquely-identified record per document. The document record has an internal structure accessible as named fields. The fields are:

- Unique Identifier (service-chosen, not changeable)
- Number of page images
- Page image (repeated field)
- Anchor map
- Full text (format to be defined)
- Text/image map (format to be defined)
- Bibliographic fields—Author, title, abstract, etc. as found in the standard bibliographic record.
- A virtual field that materializes as the standard Bibliographic record
- Deduplication value
- Transaction ID

All fields except Unique ID and Transaction ID are optional. Although the structure is that of a data-base system, the only retrieval indexes maintained by the service are for the Unique Identifier and Transaction ID fields.

Assuming that there are many such storage services, it is not yet clear whether the Unique ID should be unique only with respect to the service that assigned it, and that clients are responsible for remembering to which service to present it, or whether the service is also responsible for making that ID unique across all such services and that clients can somehow determine to which service to present it by performing some operation on the ID itself.

The anchor map is a list of pairs, consisting of named anchor points and page image indexes. Anchor points are intended to be used to indicate where in the array of page images one can find, e.g., a table of contents, a table of figures, the first page of chapter one, the conclusions, or a bibliography. Anchor point names are not interpreted by the service and not standardized; they may be provided by a conscientious service for the convenience of browsing interfaces.

The text/image map is intended to serve two purposes: a program can determine which page image and coordinate region of that image contains any word listed in the full text, and a program can identify from a coordinate region in a page image what text words lie in that region.

The deduplication value is a bitstring derived from the initial contents of the record, intended to uniquely distinguish it from every other such record. The method of computing the deduplication value is not yet defined, but it might, for example, be a 64-bit cyclic redundancy checksum of the initial value of the bibliographic record field. The deduplication value, once computed, is permanent; it is not changed even if the

thing it was computed from changes. Its purpose is to allow a client to determine that references obtained from different sources refer to the same record.

- D. The service may be implemented with replicants. Interfaces for managing replication and recovery, and guarantees about when replicants learn about committed updates are not yet defined. The feasibility of providing reliable replicants without making the service depend on the service provided by the replicants has not been established.
- E. For the use of discovery intermediaries, the storage service provides a "what's-new" interface that allows a discovery service to obtain a list of changed records. The discovery client interested in invoking the what's-new interface does not receive any notification that things have changed; the client is expected to poll. Thus clients initiate all information flows either in or out of the storage service. (One might implement an awareness service associated with the storage service, but the awareness service would also be a client that polls and makes use of the what's-new interface.)

6 Scanning System Design Issues

In a series of meetings, we made significant progress on planning and system design for document scanning. The primary result was a thorough review of the various aspects and preparation of a list of issues that need to be addressed. In parallel, Keith Glavash, Lindsay Eisan, and Jeremy Hylton prepared a complete hardware/software scanning station proposal, and that proposal is now the basis for iterative improvement.

Some of the more interesting issues uncovered are the following: Technical report images may be obtained from several sources (archive or shelf copies, guillotined cut pages, micro-film or fiche, PostScript source), and may involve any of a large number of exceptions (color, photographs, oversize sheets and foldouts, yellowing paper and fading ink). Scanning can be done at various resolutions (300, 600, 1200 pixels per inch), at various depths (black & white, gray-scale, color), and with various adjustments (brightness, contrast, linearity, threshold). In a production setting, the resulting images need to be aligned, cropped, cleaned up, quality checked, labeled and archived. Finally, each distinct prospective use of an image (storage, screen display, printing, optical character recognition) seems to require distinct processing of the image bits. This latter observation leads to many possible architectural variations involving tradeoffs between on-the-fly conversion and storage of multiple forms of the same image.

One of the single biggest problems is that the resulting data is rather voluminous, and it is easier to think up work flows that contain bottlenecks than ones that do not. The initial data from a good quality grey-scale-capable scanner consists of about 10 MBytes/page, or about 100 Kbytes/page after compression. The scanner can read 40 pages/minute, yielding 500 MBytes/minute of raw data. For comparison, SCSI-2 drives can read or write 240 MBytes/minute, DAT drives 1 MByte/minute, and Ethernet transfer typically takes place

at about 0.5 MByte/min. The implication is that one must do some form of compression very close to the scanner; unfortunately the computer scanning platform that has been most developed is the Macintosh, but that is not the easiest environment in which to develop experimental prototypes. Fortunately, a single human operator can apparently physically handle, track, coordinate, and check quality on only about 400 pages per hour, an order of magnitude less than the mechanical capability of the scanner. Thus one can reasonably consider options such as buffering a day's work on disk and transferring the resulting data dump over the network or to tape over night.

The internal memoranda of the project contain a much more detailed list of design issues that arise in the scanning and image handling processes.

7 Other, Related Activities

Mitchell Charity reviewed and brought up to the version 2.0 specification bibliographic records for about 1000 M. I. T. Laboratory for Computer Science Technical Reports. These bibliographic records are now available to the other CS-TR project participants; Stanford University has begun making use of this resource.

Jerome Saltzer attended a CS-TR planning meeting at CNRI headquarters in Reston, Virginia, on February 3, 1993. This meeting affirmed use of the RFC-1357 bibliographic record, and participants agreed to develop other standards.

8 Technology and the LCS/AI Reading Room

One of the first projects undertaken by Library 2000 was the automation of the catalog of the LCS/AI reading room, making that catalog available via internet to anyone, including, of course, the workstations in offices throughout the two laboratories. Since the online catalogue was introduced in 1991, there has been a major impact on use of the Reading Room collection. A sharp increase in the circulation of materials seems to be caused by the availability of bibliographic information both on-site and via the network. In addition, more bibliographic search tools have been implemented, e.g., CD-Barton, ACM Computing Archive, Dissertation Abstracts and INSPEC. Public-use Macintoshes in the Reading Room are now set up with double-click access to the Reading Room catalogue; FirstSearch; TechInfo; the Boston Library Consortium Gateway; and the Internet services WAIS, Archie, Gopher, and World-Wide Web, providing a 'one-stop shopping' source for the LCS/AI Community.

Maria Sensale is investigating online distribution of Technical Report Bulletins and Table of Contents Bulletins, traditionally distributed in hard-copy form.

In-room use of the collection is strong and there is a growing trend for users to regard the resource as a 'virtual library' as well. Requests for information and research support are being made by e-mail and are fulfilled online using a wide-range of document retrieval services via the Internet and commercial vendors, or by fax and traditional sources.

The M. I. T. Special Libraries Group, made up of Librarians from departmental reading rooms and libraries, as well as representatives from the M. I. T. Library System, has continued to meet on a regular basis and exchange information as part of a collegial network. The group met with the Director of the M. I. T. Library System, the Associate Director for Public Services, the Associate Director for Systems and Planning, and toured various facilities on Campus. Jerome Saltzer also gave a talk to the group about the status of Library 2000.

In order to further plans for more extensive Lab-wide out-reach, a Steering Committee is being formed to oversee Reading Room issues. In addition, the Reading Room Staff will work with Group liaisons to customize information services and to provide research support on an individual and group-wide basis.

9 Plans

The research group prepared proposals to IBM and Digital for equipment grants to carry out the CS-TR project, as well as continued development of the Library 2000 testbed. The IBM proposal (\$200K for maximum expansion of RAM memory for three existing index engine servers) was successful and we have placed an order for equipment under that grant. Recently, Digital Equipment Corporation invited us to submit a grant proposal (\$500K for triplexed 22-Gbyte storage servers), and DEC is now considering that proposal.

We intend to complete the specification of the scanning station to be installed in the MicroReproduction Laboratory, order the equipment, and beginning scanning operations over the summer. In anticipation, MRL staff have begun locating, gathering, and evaluating the quality of appearance and completeness of a comprehensive set of paper originals of LCS and AI Technical Reports. The other major scanning component still to be designed is a bitstream processing and conversion strategy, with at least four variations: for archival storage, retrieval storage, display, and printing.

If our hardware grant applications are successful, during the coming year, the data available under the Library 2000 testbed system is scheduled to expand in several ways. An update

stream will turn the M. I. T. library catalog card collection from a snapshot into a current database. We will begin to accumulate page images of M. I. T. Computer Science Technical Reports and theses. We will also begin a new collection of bibliographic records, abstracts, and pointers to the page images of the Computer Science Technical reports of the other partners in the CS-TR project. And we plan to provide a full-text index to current network news as an additional on-line collection.

The Library 2000 testbed itself is scheduled for a number of changes. First, there are a number of under-the-covers upgrades that will make prototyping faster and easier. Next, the protocols that run among the presentation client, the index server, and the storage server are ready for another round of implementation; their design in the areas of versioning, search specification, update, what's-new, authentication, and access control has progressed to the point that the ideas need to be tried out. As soon as the memory of the index servers is upgraded to 500 MBytes, the indexing system will be adapted and configured to take maximum advantage of the additional space. Similarly, availability of large storage servers will trigger work on configuring file systems for large (image) files, a new storage service protocol, integrity monitors, and a remove-to-tape facility. At the same time, it should be possible to try out some initial ideas for naming of collections, storage services, index services, and documents and provide a first implementation of the link from a bibliographic record to a document image.

10 Publications, Theses, and Talks

Published papers:

Anderson, Greg (with a sidebar by Lucker, Jay K.), "Mens et Manus at Work: The Distributed Library Initiative at MIT," in *Library Hi-Tech* 11:1 (1993, Issue 41) pp 83-94.

Saltzer, J. H., "Technology, Networks, and the Library of the Year 2000," in *Future Tendencies in Computer Science, Control, and Applied Mathematics, Lecture Notes in Computer Science 653*, edited by A. Bensoussan and J.-P. Verjus, Springer-Verlag, New York, 1992, pp. 51-67. (*Proceedings of the International Conference on the Occasion of the 25th Anniversary of Institut National de Recherche en Informatique et Automatique (INRIA)*, Paris, France, December, 1992.)

Saltzer, J. H., "Needed: A Systematic Structuring Paradigm for Distributed Data," *Operating Systems Review* 27, 2 (April, 1993), pp. 77-81. Originally distributed as paper #41 in *5th ACM SIGOPS Workshop on Models and Paradigms for Distributed Systems Structuring*,

September 21-23, 1992, Le Mont Saint-Michel, France, pp. 1-5.

Gong, Li, Lomas, T. Mark. A., Needham, Roger, and Saltzer, Jerome H., "Protecting Poorly Chosen Secrets from Guessing Attacks," scheduled to appear in *IEEE Journal on Selected Areas in Communications* 11, 5, June, 1993, pp. 1-9.

Talks:

Saltzer, J. H. "Technology, Networks, and the Library of the Future." IBM Research Laboratory, Hawthorne, New York, February 26, 1993.

Saltzer, J. H. "Storage, the Unnoticed Revolution." Open session of the Computer Science & Telecommunications Board, National Research Council, Washington, D. C., May 24, 1993.

Anderson, T. Gregory. "Electronic Publishing." Harvard Library Automation Planning Committee, Harvard University, Cambridge, Massachusetts, June 4, 1993.