Title
KEYWORD ACCESS TO A MASS STORAGE DEVICE AT THE RECORD LEVEL

Author
Gey, Fredric

Publication Date
1975-09-01

To be presented at the International
Conference on Very Large Data Bases,
Boston, MA, September 22 - 24, 1975

LBL-4256

KEYWORD ACCESS TO A MASS STORAGE DEVICE AT THE RECORD LEVEL

Fredric Gey and Marilyn Mantei

September 12, 1975

Prepared for the U. S. Energy Research and
Development Administration under Contract W-7405-ENG-48

DISCLAIMER

KEYWORD ACCESS TO A MASS STORAGE DEVICE AT THE RECORD LEVEL

BY

FREDRIC GEY AND MARILYN MANTEI

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94709

Abstract

A general software package was built to access individual data records stored on an IBM-1360 photo-digital mass storage device with a current on-line capacity of 50 billion characters (bytes) and an infinitely extensible off-line capacity. An existing data base management system was used to maintain the pointers to the data on the mass storage device and to store the controls for the data driven interactive code. Existing data dictionaries used for sequentially processing the data bases were stored in the DBMS and used to display individual data items within the retrieved records.

Data retrieved from the mass storage device is displayed interactively or sent for further processing to various report generation and statistics packages. The system provides dial-up terminal retrieval capability for exceedingly large socio-economic and demographic data bases used by national and regional planning agencies of the federal government. It was developed with three man-months of effort.

This work was supported by the U.S. Energy Research and Development Administration with funding from the U.S. Army Corps of Engineers (St. Louis District and Office of the Chief of Engineers) and U.S. Dept. of Labor (Manpower Administration).

Authors' addresses Fredric Gey, Mathematics and Computing Department, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720; Marilyn Mantei, The Annenberg School of Communication, University of Southern California, Los Angeles, California 90007.

1. INTRODUCTION

The data access system described in this paper was built with the following constraints placed upon its construction.

1. It had to be built in a relatively short time frame, i.e., three months.
2. It had to access extremely large socio-economic-demographic data bases (whose potential size exceeded a single disk pack).
3. The access required was not totally determinable, i.e.,the users could not predict the data requests they would make.
4. It had to have an interactive capability with a language designed for the casual (non-programmer) computer user.

These goals were met by utilizing the hardware available at Lawrence Berkeley Laboratory and by managing that hardware using a general data base management system. The project was staffed by a single programmer who had no previous knowledge of the DBMS software he utilized. With the DBMS and the ability to modify job control streams on the laboratory's CDC computers, a software system was put into operation on February 1, 1975 after three months of programming effort.

The data base management system used was an in-house product built to handle another large on-going project. It is called STOFI (1), an acronym for storage file, and is essentially a set of FORTRAN callable routines to provide index sequential and hierarchical access to files stored on disk.

The data to be accessed was stored as several files on an on-line photo-digital mass storage device (chipstore) [2], [3], [4] capable of handling fifty billion characters of data. The chipstore hardware allows data transfers of a minimum size of 20 CDC 60-bit words. With this hardware capability and the STOFI system, a directory of chipstore addresses was built in a STOFI type file. Keys were selected from the data record and multiple access paths were incorporated into the directory file.

The system currently accesses data from the U.S. Census of Agriculture (84 million characters), the City-County Data Book time series (20 million characters), the BEA Income time series (20 million characters), the 1960 U.S. Census of Population (40 million characters) and the series C and E OBERS Projections of the U. S. Water Resources Council (40 million characters). Plans are to extend the system to cover files now on chipstore and in our tape library that exceed 1 billion characters.

The first part of this paper describes in detail the environment under which the system was built. Section 2 explains the BKY operating system and the mass storage device. Section 3 lists the features of the software borrowed to implement the project.

Section 4 describes the access system and details the structure and information characteristics of the data bases being accessed, and sections 5 through 7 list the project's operational characteristics including costs, speed, drawbacks, and future plans. A summary at the end describes the general nature of the data access system.

2. HARDWARE AND OPERATING SYSTEM ENVIRONMENT

THE LBL COMPUTER NETWORK

The hardware environment of the Lawrence Berkeley Laboratory computer facility consists of a network of 3 computers, a CDC 6600 with 132K words of central memory, one CDC 6400 with 64K of central memory and one CDC 7600 with 64K of central memory and 512K of large core memory. [Figure 1]

Both the 6000 series machines serve as feeders to the CDC 7600. In addition, the CDC 6000 series machines operate under a locally built batch multiprogramming system (BKY) [5], [6] which allows for an interactive interface via a PDP-8 system called RECC (for Remote Equipment Control Computer). At present 160 consoles are handled under this system. Both 6000 machines can multi-process up to 64 jobs (control points) at any one time. Only the 6600 machine has an access channel built to the photo-digital storage device but data can be staged to the other computers.

All data manipulation done by individual user programs is carried out on random access disk files. Files are staged to these disks from permanent storage and saved after program operation at the discretion of the user. Available permanent storage consists of 4 IBM-2321 data cells, a 40,000 reel tape library, and the IBM-1360 photo-digital storage device.

A special card-image file called CONTROL (for each control point) contains the job control stream for excecuting system commands. The system steps through this file, one command at a time. Changes to the job flow occur if errors are detected or if the file is rewritten with a new set of job control cards. The data access system being described uses this feature to fetch its permanent files from chipstore once it has been determined via the interactive session what data is needed.

THE MASS STORAGE DEVICE

The IBM-1360 photo-digital storage system is a peripheral storage device composed of a storage file containing 2250 boxes of silver halide film chips, a chip recorder-developer, and a chip reader. Figure 2 illustrates its general hardware arrangement and its connection to the CDC 6600 computer.

Each chip holds 4.7 million bits of data as well as error correction and detection and addressing codes. Up to 32 chips are stored in a single box. These chips are recorded with an electron beam and read via a photo-multiplier tube detection scheme. The total system holds up to 5 x 10**10 characters of data on-line. In addition, off-line storage requiring manual retrieval extends the storage capacity of this device infinitely. It is possible to build a data base in this system equivalent to 1000 reels of 9-track tape.

3. EXISTING SOFTWARE PACKAGES

THE MSS SYSTEM

A mass storage system (MSS) has been written at LBL to manage chipstore. Typical access is by volume and data set requests. However, it is possible to retrieve data by specifying the chipstore address of the data desired. Since each chip is encoded with addressing information, it is possible to create a sequential file of records in the non-procedural MSS control language and read that same sequential file in a lower level MSS procedural language to obtain the address locations for each record in the file. It is these addresses that are then stored as the file directory in the STOFI system.

THE STOFI PACKAGE

The STOFI package is a series of FORTRAN callable subroutines that allow a user to multi-key his data and build hierarchies of data within his data base. The data access scheme consists of directories and sub-directories which are interrogated with a binary search algorithm. Directories and data are interleaved on the user's file and page buffering and I/O optimization are managed by the STOFI system.

Keys, records and data fields can all be of variable length, and the dynamic work space used by STOFI is available to the user. A STOFI feature used quite often is the ability to build directories (index files) within directories to any depth. This feature and the capability of multi-keying a record allow a user to build many logical access paths to his data.

The STOFI system was built to efficiently manage large disk-based data bases. Since a large directory was expected for the chipstore data bases, this system was chosen over another in-house DBMS built for smaller data bases. STOFI typically manages 20 million byte files but is working effectively on a 100 million byte file. Ten basic subroutine calls were needed to build the large scale data access system around STOFI.

THE QWICK QWERY PACKAGE

QWICK QWERY [7] is a proprietary retrieval and report generation system developed by C.A.C.I., Inc. It is batch oriented and works on fixed length records and fields. It is operational only on the CDC-7600 computer and requires a user defined dictionary to access the records it processes. Simple aggregations, statistics, sorts and boolean selections are available with the report formatting. The QWICK QWERY system operates by making a sequential pass over the entire data base.

The QWICK QWERY system had been used to prepare reports from several of the large data bases which were to be accessed interactively with the system this paper describes. Since dictionaries had been created which described the individual fields in each record of these data bases, these same dictionaries were stored as data in the STOFI system. The individual data definitions were then used to display the retrieved data during an interactive session.

4. DESCRIPTION OF THE SYSTEM

The system operates in three parts as illustrated in Figure 3. Initially the file or files are stored as data sets on chipstore. Once stored, they are read sequentially and a directory to the records is built based on a preprogrammed hierarchy set up by the user. The QWICK QWERY dictionary, the interactive help commands and the various job control streams are also built into the directory file at this time. The third portion of the system is the interactive retrieval program, called REAP [8]. REAP fetches entire records or groups of records for the user and displays requested portions of them on the user's console.

Figure 4 is a run-time scenario using REAP to access the City-County Data Book files. Note that data is retrieved into a temporary file (cache) as displays are requested. The cache is retained until a new one is requested from the mass storage device. Caches may be merged or subsets of them used for varied displays. Figure 5 lists the language commands needed to operate the interactive program.

Each data base may have one or many logical structurings. These may be built into a single STOFI file or into many STOFI files depending upon the user's needs. A job control stream is set up and executed from user requests to bring in the desired directory file.

The data access paths defined in a given file may be changed as needed. Controls are built within the interactive program to use the file's logical structure to monitor the user's search such that the access to the sequential data base on chipstore can be modified at will with little cost to the user or programmer. Figure 6 shows the file structure being used to access the City-County Data Book and OBERS Projections by regional planners. The access paths are geographic in content because the data base's current users wish to work within their geographic area. Note that data describing the data base's contents is stored on the directory file. This data plus data describing the structure of the file can be displayed by using the ∧HELP and ∧LIST commands in the REAP language.

The cached data can be fetched via a batch or interactive request and sent to a report generation or graphics display package. This larger volume work is routed to the CDC 7600 and then to remote users. The job control streams stored within the directory file may be used to direct this routing. See Figure 7 for a sample report generated from a cached file built by merging several data sets of the U.S. Census of Agriculture data base.

5. PROJECT COSTS

Development costs include nominal machine debug time and three man-months of programmer effort. Costs were kept low by employing existing data management software and an existing data description dictionary.

No significant processing time overhead resulted from adding another software package on top of the MSS software. Considerable savings were achieved by not searching the entire sequential file for display data. Figure 8 presents costs for two sample report runs. Run 1 entails displaying a small amount of data selected from two large chipstore data sets. Run 2 presents access time only for a large volume of data from a single chipstore data set. The QWICK QWERY batch runs were carried out by staging the entire chipstore data sets to the CDC 7600 for QWICK QWERY processing.

A sufficient amount of performance tests have not been carried out for the cost comparison of both systems. There are inherent difficulties in this comparison because the interactive and batch programs run on two different machines which possess different operating systems and I/O buffering characteristics. Thus a change in the accounting algorithms could greatly change Figure 8. It is, however, apparent that a major saving in running REAP is gained from the ease of access the user has to his large data bank.

6. DRAWBACKS OF THE KEYWORD ACCESS SYSTEM

The primary limitation of this data access system is the computing environment at the Lawrence Berkeley Laboratory. Programs, to run interactively, must be 15-20K in size. The REAP program currently runs in 30K of small core memory. This causes it to reside on disk for more than a comfortable wait period until a slot of memory becomes available. Cures in the form of segment linking and placing more of REAP's control data on the directory file are being implemented.

Long maintenance periods (4-8 hours per week at prime run-time) for the chipstore are a second drawback to running the interactive access system. Programming is being done now to incorporate the tape library and the data cell into this system so that data may be stored and accessed from a variety of devices. Currently, fixed length fields and records are required on the data base. This is historical dating from the requirements of the QWICK QWERY package. The package has no update capabilities. Data is written on chipstore in sequential files. If the files require changing, they must be rewritten. It is, however, possible to add data to the data base by creating additional chipstore data sets and linking these to the STOFI directory.

Finally, the user is required to have a basic knowledge of the data in order to use the system. Although help options are sprinkled liberally throughout the REAP program, a user still must have a basic concept of the logical structuring put on the data with his directory file. He also requires a printed dictionary of the data elements before him when running since the dictionary is too large for interactive listing and no dictionary search software has been built.

8. FUTURE PLANS FOR THE SYSTEM

The first programming planned for improving the system is the generalization of the directory build. Currently, a one-shot program is required to build the index file structure. Interactive editing of the job control streams and the QWICK QWERY dictionary is already a reality, and plans to allow building and modifying the data access structure interactively are being made.

Data is retrieved into a disk file from MSS and discarded when a run is completed. It is often desirable to save a current cache for a given week's or month's processing. This semi-permanent cache system is expected to be put into the data access scheme by job control stream modifications.

The ability to handle variable length records and fields within a data set is also being planned. Report generation capabilities are being expanded and a more general system involving modeling and statistical software application to the cache retrieved from chipstore are planned. Most of these additions will be implemented again by the interfacing of already available software.

The final plan of the access system is to generalize the access to magnetic tape and data cell data sets such that a request for data will be transparent to the storage media. It is desired that the access language will also work on a new, larger mass storage device to be purchased by LBL.

9. SUMMARY

An access package was developed to content address a large on-line mass storage device via an interactive session. The design of this package involved interfacing two existing software packages and predefined dictionary files for the data bases. The project took three man-months to complete.

The salient features of the system are

1. Data bases are stored in sequential order. The logical file structure is maintained independent of the actual data.

2. The logical structure is a hierarchy of keys taken from the data element values within each record. Keys may be specified either by name (taken from the data dictionary for the file), by position (location within the record) or by user supplied values.

3. Data dictionaries are stored on the index file and used to display individual fields in a record.

4. The interactive code is data driven such that, help commands and promptings will depend on the structure of the user's file directory.

5. Help aids, user lists and job control streams are stored on the user's directory file.

6. The data is archival in nature, i.e., the system is a retrieval not an update system.

References

1. Mantei, M. and Sarna, J., An introductory manual to STOFI, LBL-2649, LBL, Berkeley, CA (1974).

2. Penny, S.J., Fink, R. and Alston-Garnjost, M., Design of a very large storage system, Proc. of AFIPS FJCC, 37 (1970) 45-51.

3. Kuehler, J.D. and Kerby, H.R., A photo-digital mass storage system, Proc. of AFIPS FJCC, 29 (1966) 735-742.

4. Furman, R.M., IBM 1360 photo-digital mass storage system, IBM technical report, Tr 02.427 (may 15,1968).

5. Stevens, D.F., On overcoming high priority paralysis in multiprogramming systems, a case history, COMM. ACM 11, 8 (1968) 539-541.

6. Tannenbaum, A.S. and Benson, W.H., The people's time sharing system, Soft.-Prac. and Exp., 3 (1973) 109-119.

7. CACI Inc., QWICK QWERY user's manual, Consolidated Analysis Centers, Inc., 12011 San Vicente Blvd., Los Angeles, CA 90049.

8. Gey, F. and Williams, E., REAP, a teletype retrieval program for the socio-economic-environmental-demographic information system (SEEDIS), UC.ID. 3751, LBL, Berkeley, CA (1975).

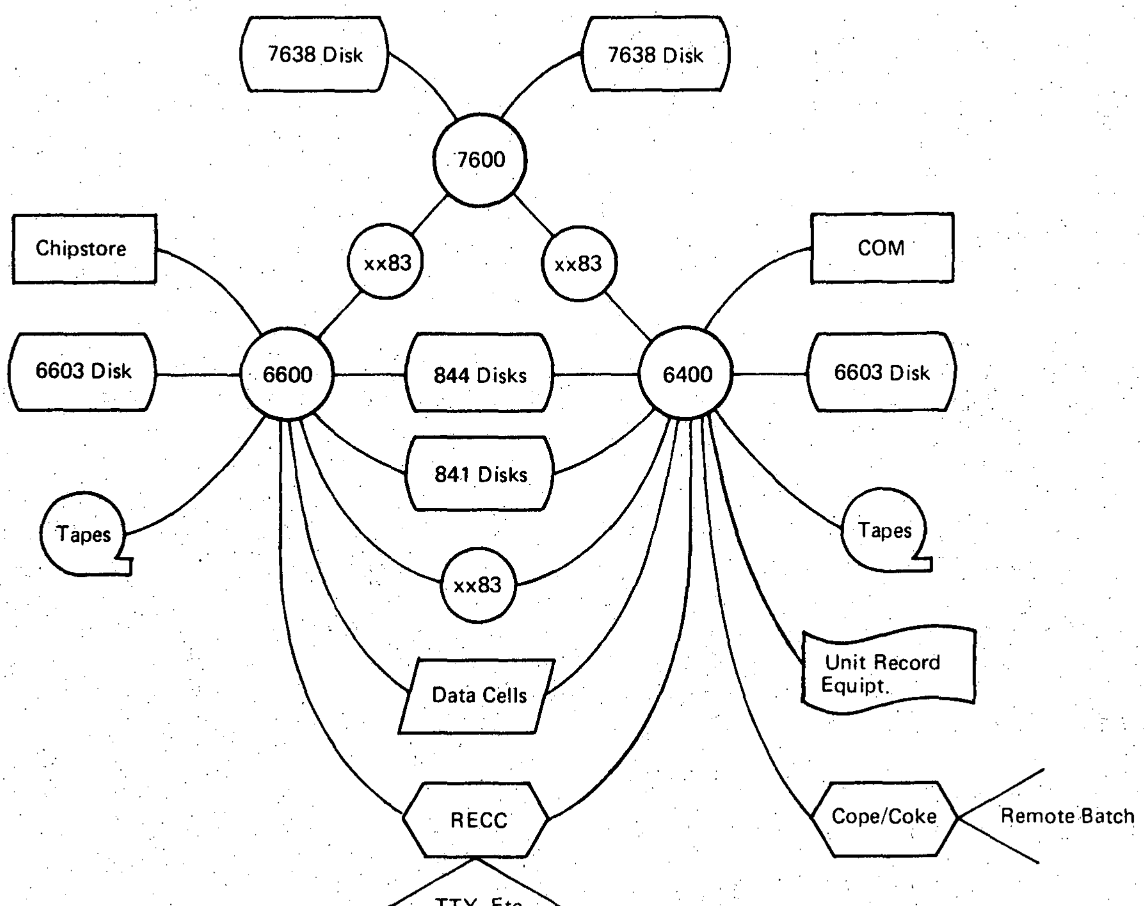

Figure 1. The LBL Computer Center

XBL 758-3721

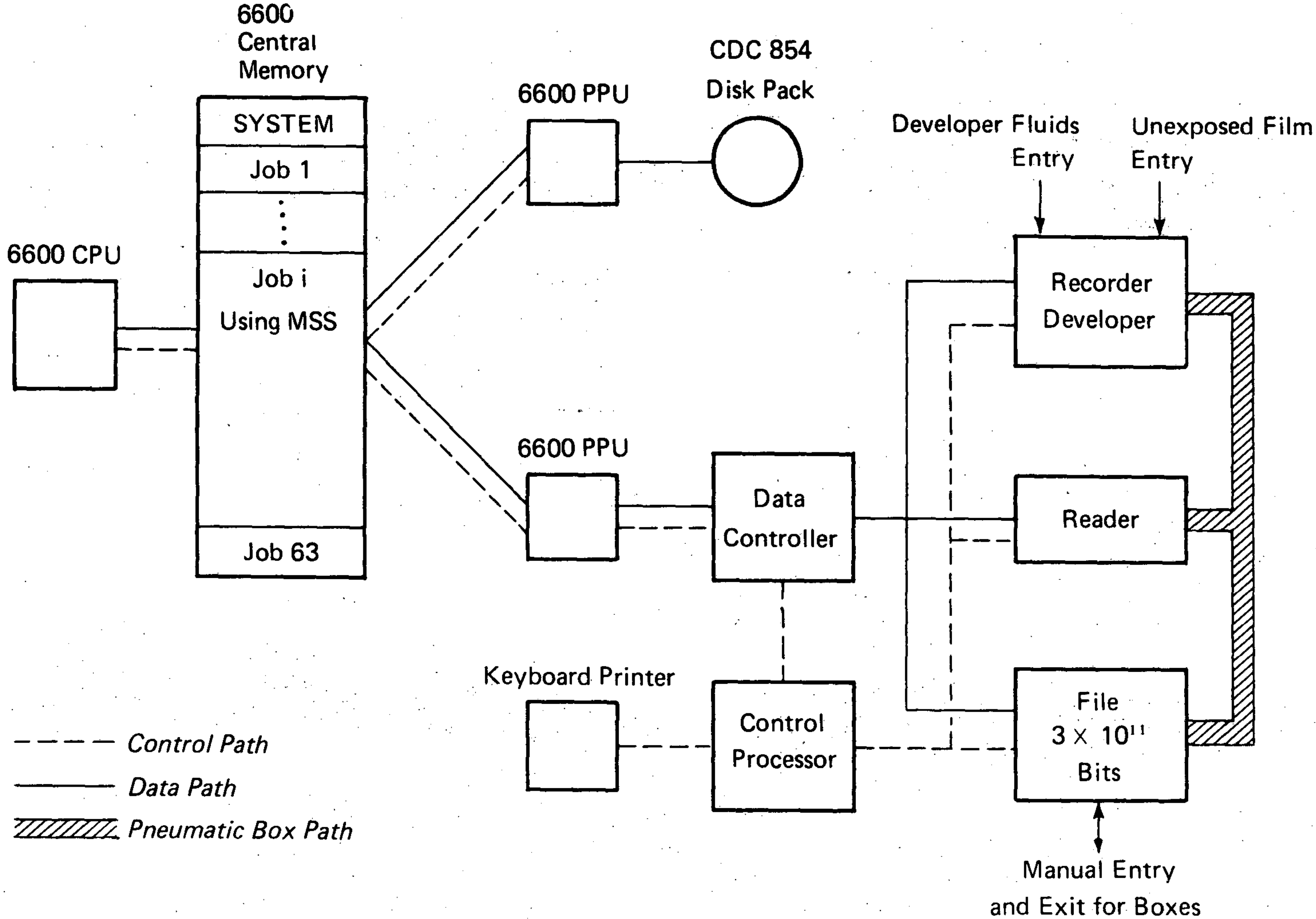

Figure 2. General MSS Architecture

XBL 758-3720

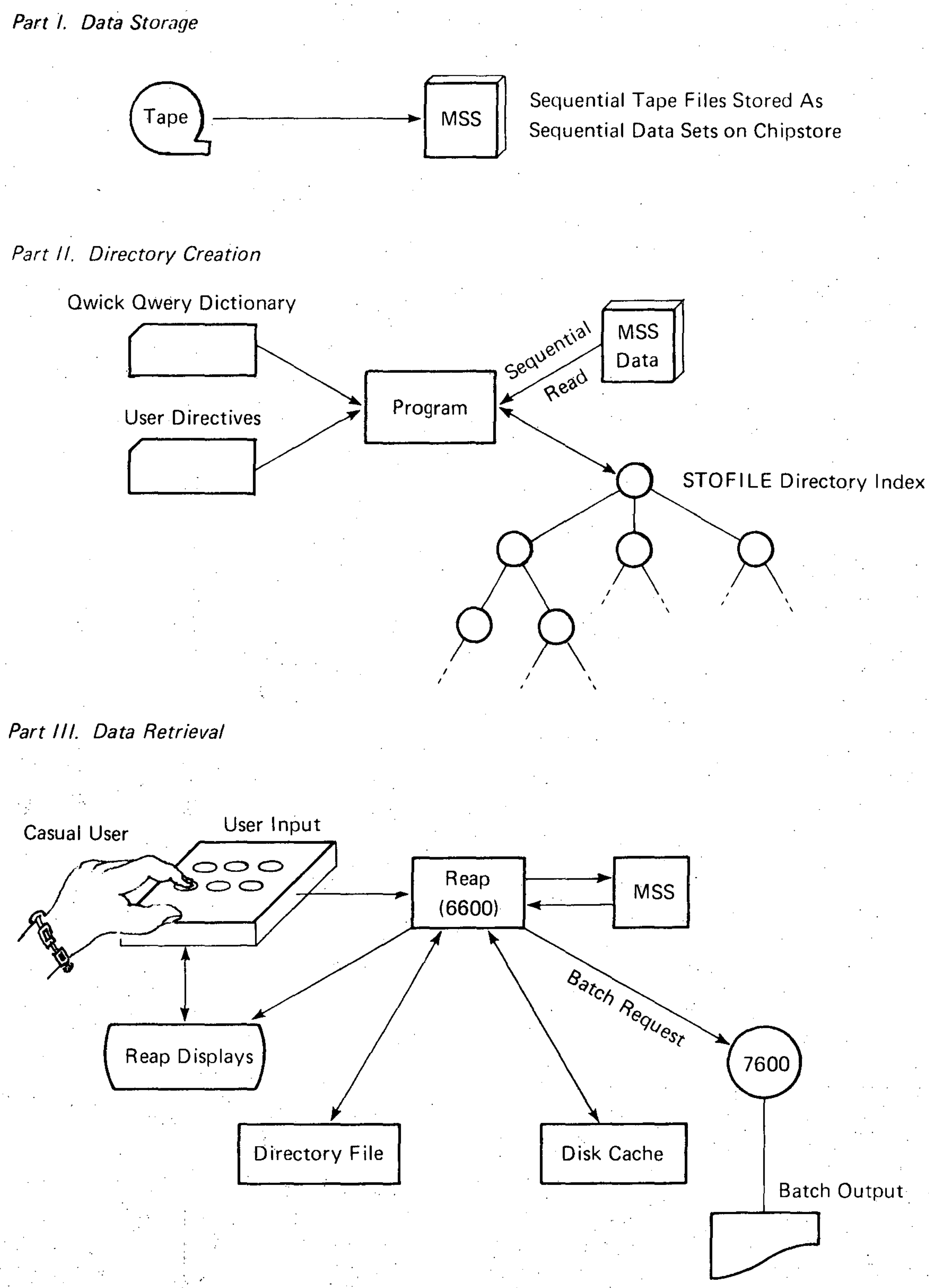

Figure 3. Tripartite System Operation

XBL 758-3719

```
WELCOME TO THE LBL
 - - S - E - E - D - - -
SOCIO-ECONOMIC-ENVIRONMENTAL-DEMOGRAPHIC
INFORMATION SYSTEM RETRIEVAL PROGRAM
ENTER FILE NAME
AHELP
  LIST OF AVAILABLE FILES

LBL-NAME  QWICK QWERY NAME   DESCRIPTION

CCDBMRG   CCDB.MERGE         CITY-COUNTY DATA BOOK (1952-1972)
CNAGR49   49.AGR.SUB         1949 U. S. CENSUS OF AGRICULTURE
CNAGR59   59.AGR.SUB         1959 U. S. CENSUS OF AGRICULTURE
CNAGR64   64.AGR.SUB         1964 U. S. CENSUS OF AGRICULTURE
CNAGR69   69.AG.CENCOM       1969+1964 U. S. CENSUS OF AGRICULTURE
CTY60PC   COUNTY.60.PC       1960 CENSUS OF POPULATION
INCOMER   INCOME-REV         REVISED BEA INCOME TIME SERIES 1929-1969
OBERS     OBERS              SERIES C OBERS PROJECTIONS 1970
OBERSE    OBERS-E            SERIES E OBERS PROJECTIONS 1972
ENTER FILE NAME
CCDBMRG
 FILE=CCDBMRG-CCDB.MERGE     CITY-COUNTY DATA BOOK (1952-1972)
ENTER CATEGORY NUMBER
1
DEFINE STUDY AREA
NEW OR OLD?
NEW
ENTER STUDY AREA NAME
FGTEST
ENTER AREA CODES
17013,17083,17149;
DISPLAY?
STATE.ABBR,COUNTY.NAME,52.CCDB048,52.CCDB049,62.CCDB051,72.CCDB077;
OUTPUT LINE LENGTH FULL
RETRIEVING DATA FROM MSS, PLEASE BE PATIENT
 .
 .
 .
 STATE    COUNTY                  1940     1950     1960     1970
 ABBR      NAME                DWELLING DWELLING  HOUSING  HOUSING
                                 UNITS    UNITS    UNITS    UNITS
ILL     CALHOUN                   2325     2340     2531     2302
ILL     JERSEY                    3939     4895     6003     6586
ILL     PIKE                      7822     7913     7928     7790

*TOTALS*                       -------------------------------------

                                 14086    15148    16462    16678
MORE QUERIES ON SAME STUDY AREA?
ASTOP
END-OF-JOB
```

Figure 4. A REAP Scenario

I REAP GLOBAL COMMAND - viable responses to all REAP queries

↑ HELP - provide help messages
↑ LIST - list user input to current task
↑ DISCARD - restart current task discarding previous input
↑ QUIT - stop current processing; begin again at start of prog.
↑ STOP - end all processing (normal exit procedure)
↑ ABORT - end all processing (abnormal exit procedure)

II REAP CONTROL LANGUAGE

1 - ENTER FILE NAME - Enter chipstore data set name. If unknown, ask for help.

2 - ENTER CATEGORY NUMBER - Enter search path desired for data access. If unknown ask for help.

3 - DEFINE STUDY AREA - Enter 'OLD' or 'NEW' to indicate current cache or new cache usage.

4 - ENTER STUDY AREA NAME - Assign label to cache.

5 - ENTER AREA CODES - List keys for data desired to be fetched from data set. A semicolon ends the entry process.

6 - DISPLAY - List the data elements desired for display. A semicolon ends the entry process.

7 - MORE QUERIES ON THE SAME STUDY AREA - A 'YES' will allow additional displays. A 'NO' will return to step 1.

Figure 5. REAP interactive query language

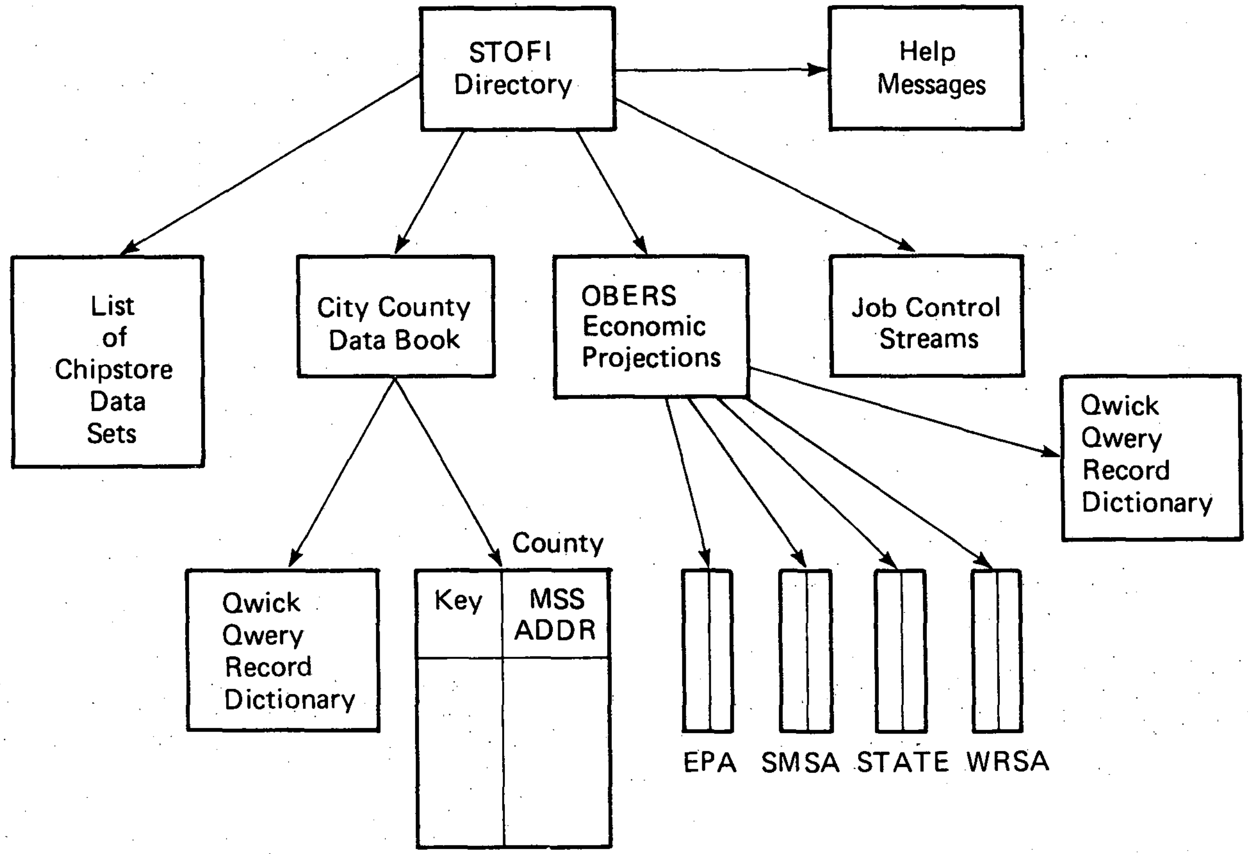

Figure 6. Sample STOFI Directory file for US Economic Data
(City-County Data Book and OBERS Projections)

XBL 758-3722

U.S. CENSUS
OF AGRICULTURE
DATA FOR THE YEARS
1949,1959,1964,1969

UPPER MISSISSIPPI RIVER
SOCIO-ECONOMIC STUDY
MINNEAPOLIS-ST.PAUL SMSA
LOCK/DAM 26 IMPACT STUDY

LAWRENCE BERKELEY
LABORATORY
UNIV. OF CALIFORNIA
13 AUG 75

	1949	1959	1964	1969
TOTAL NUMBER OF FARMS	21,474	17,042	15,170	11,728
DECLINE DUE TO DEFINITION	-NA-	452	0	-NA-
TOTAL LAND AREA	3,307,520	3,307,520	3,307,520	2,973,440
ACRES IN FARMS	2,862,110	2,616,530	2,515,399	1,860,341
ACRES IN POULTRY FARMS	-NA-	-NA-	221	-NA-
ACRES IN DAIRY FARMS	-NA-	-NA-	7,156	-NA-
VALUE OF FARM PRODUCTS SOLD **				
ALL FARM PRODUCTS SOLD	104,898	129,649	130,801	-NA-
ALL AGRICULTURAL PRODS. SOLD	-NA-	-NA-	120,770	150,101
ALL CROPS SOLD	14,552	27,720	26,728	-NA-
CROPS INCL. NURSERY PRODS + HAY	-NA-	-NA-	26,064	31,364
FIELD CROPS				
EXCLUDING VEG.,FRUITS,NUTS	9,523	20,858	19,835	-NA-
VEGETABLES	2,011	2,296	2,270	-NA-
FRUITS AND NUTS	464	452	398	-NA-
FOREST+HORTICULTURAL PRODS.	2,647	4,114	4,222	-NA
FOREST PRODUCTS	-NA-	-NA-	-NA-	93
LIVESTOCK+LIVESTOCK PRODUCTS	90,254	101,929	103,899	-NA-
POULTRY+POULTRY PRODUCTS	13,709	11,131	11,048	-NA-
LIVESTOCK+POULTRY+THEIR PRODS	-NA-	-NA-	94,401	118,645
LAND IN FARMS BY USE(ACRES)				
ACRES IN COMMERCIAL FARMS	-NA-	-NA-	1,811,708	1,557,957
CROPLAND, HARVESTED *	1,609,597	1,482,187	1,343,274	856,118
USED FOR PASTURE ONLY *	193,288	149,775	133,834	154,058
NOT HARVESTED OR PASTURED	72,113	111,455	234,832	-NA-
IN COVER CROPS *	-NA-	-NA-	-NA-	47,435
WOODLAND PASTURED *	377,418	286,314	249,635	118,223
WOODLAND NOT PASTURED *	75,360	106,098	101,211	58,018
TOTAL WOODLAND	-NA-	-NA-	267,510	176,241
COMMERICAL FERTILIZER-ACRES *	-NA-	-NA-	-NA-	390,041
COMMERICAL FERTILIZER-TONS *	-NA-	-NA-	-NA-	58,852
USE OF LIME - ACRES *	-NA-	44,634	36,779	17,715
USE OF LIME - TONS *	-NA-	112,517	97,504	43,572

*N.B. THE DATA FOR 1969 APPLIES TO COMMERCIAL FARMS ONLY
** IN THOUSANDS OF DOLLARS

Figure 7. Sample Report From a Cache File

	Interactive	*Batch*
Runs	*Reap*	*Qwick Qwery*
1. Complex retrieval of a few records from 2 MSS data sets	$2.83	$16.71
2. Access only of large cache from single MSS data set	$1.56	$ 2.10

Figure 8. Cost Comparison — Interactive Indexed versus Batch-Sequential

XBL 758-3723

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720