

ARCHIVING ONLINE DATA TO OPTICAL DISK

By J. L. Porter, J. L. Kiesler, and D. A. Stedfast

U.S. GEOLOGICAL SURVEY

Open-File Report 90-575



Reston, Virginia
1990

U.S. DEPARTMENT OF THE INTERIOR

MANUEL LUJAN, JR., Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

Chief, Distributed Information System
U.S. Geological Survey
Mail Stop 445
12201 Sunrise Valley Drive
Reston, Virginia 22092

Copies of this report can be
purchased from:

U.S. Geological Survey
Books and Open-File Reports Section
Federal Center, Bldg. 810
Box 25425
Denver, Colorado 80225

CONTENTS

	Page
Abstract	1
Introduction	2
Types of optical storage	2
Storage media costs and alternative media used for data archival.	3
Comparisons of storage media	3
Magnetic compared to optical media	3
Compact disk read-only memory compared to write-once/read many media	6
Erasable compared to write-once/read many media	7
Paper and microfiche compared to optical media	8
Advantages of write-once/read-many optical storage	8
Archival procedure and results	9
Summary	13
References	15

ILLUSTRATIONS

Figure 1. Costs of alternative forms of data storage	4
Figure 2. Average annual long-term storage cost per 100 megabytes of data	5
Figure 3. Sample of unit value archival data file	11
Figure 4. Sample of unit value index file	12

ARCHIVING ONLINE DATA TO OPTICAL DISK

By J.L. Porter¹, J.L. Kiesler², and D.A. Stedfast³

ABSTRACT

The U.S. Geological Survey has stored instantaneous values of hydrologic data (unit values) on minicomputers since 1985. A substantial amount of disk storage is required for the online storage of these data. Traditionally, these data have been archived on magnetic tape to make disk storage space available for additional data. However, magnetic tapes have a limited shelf life, and retrieval of data for a specific site is cumbersome. As the volume of unit-value data to be archived has expanded, the need for a more efficient method for data storage and retrieval has increased.

The U.S. Geological Survey's Distributed Information System Program Office is currently (1990) assessing optical disk storage as an alternative means of archiving unit-value data. Optical disks have a longer shelf life than magnetic tapes, and retrieval of archived data is substantially easier. The cost for data storage on write-once/read-many optical disks is comparable to that of magnetic tape and only a fraction of the cost for fixed magnetic disk (hard disk) data storage.

In a test study, unit-value data was archived using one optical disk drive connected to a microcomputer and a second drive connected to a minicomputer. The cost of data storage, alternative forms of data storage, and specifications for data retrieval are described.

1U.S. Geological Survey, Reston, Va.
2U.S. Geological Survey, Tampa, Fl.
3U.S. Geological Survey, Urbana, Il.

INTRODUCTION

U.S. Geological Survey (USGS) accumulates hydrologic data that require permanent archival. These records can be in electronic, paper, or other forms. The USGS conducted a study to compare archival systems and evaluate the feasibility of using optical storage media to archive hydrologic data.

To evaluate the feasibility of using optical media for archiving hydrologic data, unit-value data from a USGS office was archived on both a microcomputer and minicomputer optical storage system. This paper (1) describes types of optical storage, (2) illustrates information on storage media costs and alternative forms of data storage, (3) compares paper, microfiche, write-once/read-many (WORM), compact disk read-only memory (CD ROM), and magnetic tape and fixed magnetic disk (hard disk) media, (4) documents benefits of optical storage, and (5) presents the archival procedures and results.

TYPES OF OPTICAL STORAGE

Current optical storage devices can be classified into one of three categories according to the permanence of the data they store--CD ROM, WORM, and rewritable or erasable (Owen, 1989):

1. Read-only media such as the audio compact disk (CD) and CD ROM disks are fabricated by a pressing process, like phonograph records, where data are permanently embedded onto the disk. These disks are used to distribute large volumes of data, which can neither be altered nor erased.
2. WORM disks are supplied with no information written on the surface. The write-once disk allows the user to write data to the disk, where it cannot be altered or erased, but can be read many times.
3. Rewritable optical disks can be written to, read from, altered, and erased. Because rewritable optical disks can be altered, they are inappropriate for applications involving data archiving.

STORAGE MEDIA COSTS AND ALTERNATIVE MEDIA USED FOR DATA ARCHIVAL

The costs of alternative forms of various media required to archive 100 MB (megabytes) of information are shown in figure 1. These costs are for media only and do not include any processing or related equipment costs. The most expensive medium is paper, costing about \$2,400, followed by magnetic disk, costing about \$2,000. The least expensive is 8-mm (millimeter) tape cartridge, costing about \$7. The three least costly media for archiving data are magnetic tape, optical disk, and 8-mm tape cartridge.

The average long-term cost of storing the media required to archive 100 MB is shown in figure 2. Paper, by far the most expensive, costs approximately \$175 per year to store 100 MB. To store 100 MB of data on an 8-mm tape cartridge costs less than \$0.01. The cost of storing on write-once optical media is comparable to that for 8-mm tape cartridges.

Another storage method is the Bernoulli Box¹, with removable cartridges. One 5 1/4-inch Bernoulli cartridge drive holding 20 MB costs about \$2,500; each data cartridge costs about \$100. Even these units require 5 to 20 cartridges to store as much as one optical cartridge holds (Bican, 1988).

COMPARISONS OF STORAGE MEDIA

Magnetic Compared to Optical Media

Magnetic media have the advantage that they can be edited readily, erased, and reused. However, magnetic tape has the disadvantage that it cannot be accessed randomly like magnetic disks and optical media. Magnetic media are easily damaged by static electricity, stray magnetic fields, and excessive temperatures or humidity and, therefore, are not acceptable archival media. Access time for optical media is not as fast as for magnetic disks but is somewhat faster than magnetic tape drives. Optical media have the following advantages over magnetic media (Shier, 1989):

¹ Use of trade names or trademarks in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

Various ways of storing 100 million characters of information (100 megabytes)

Modified from:
Information Storage, Inc.

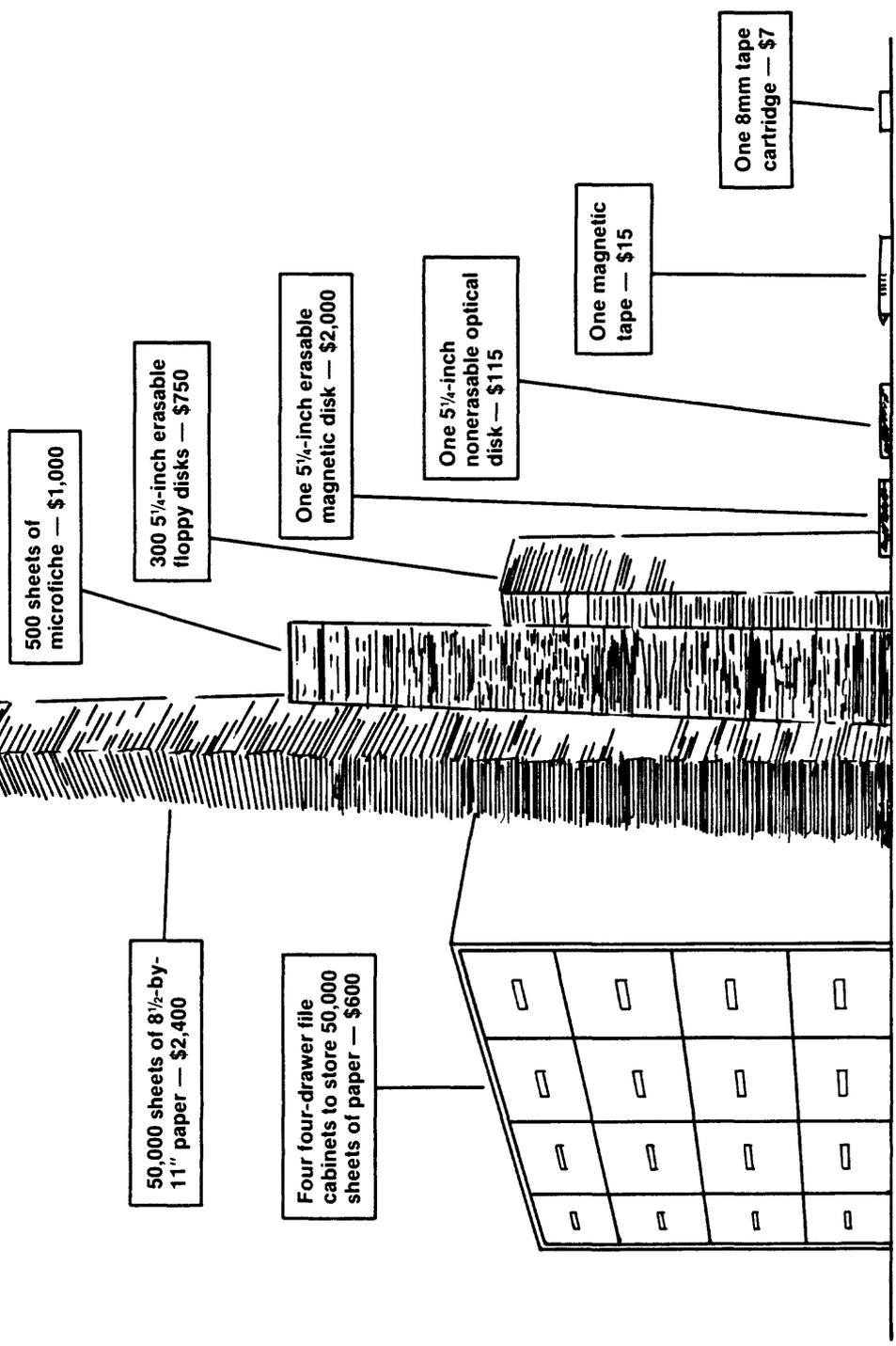


Figure 1. Costs of alternative forms of data storage.

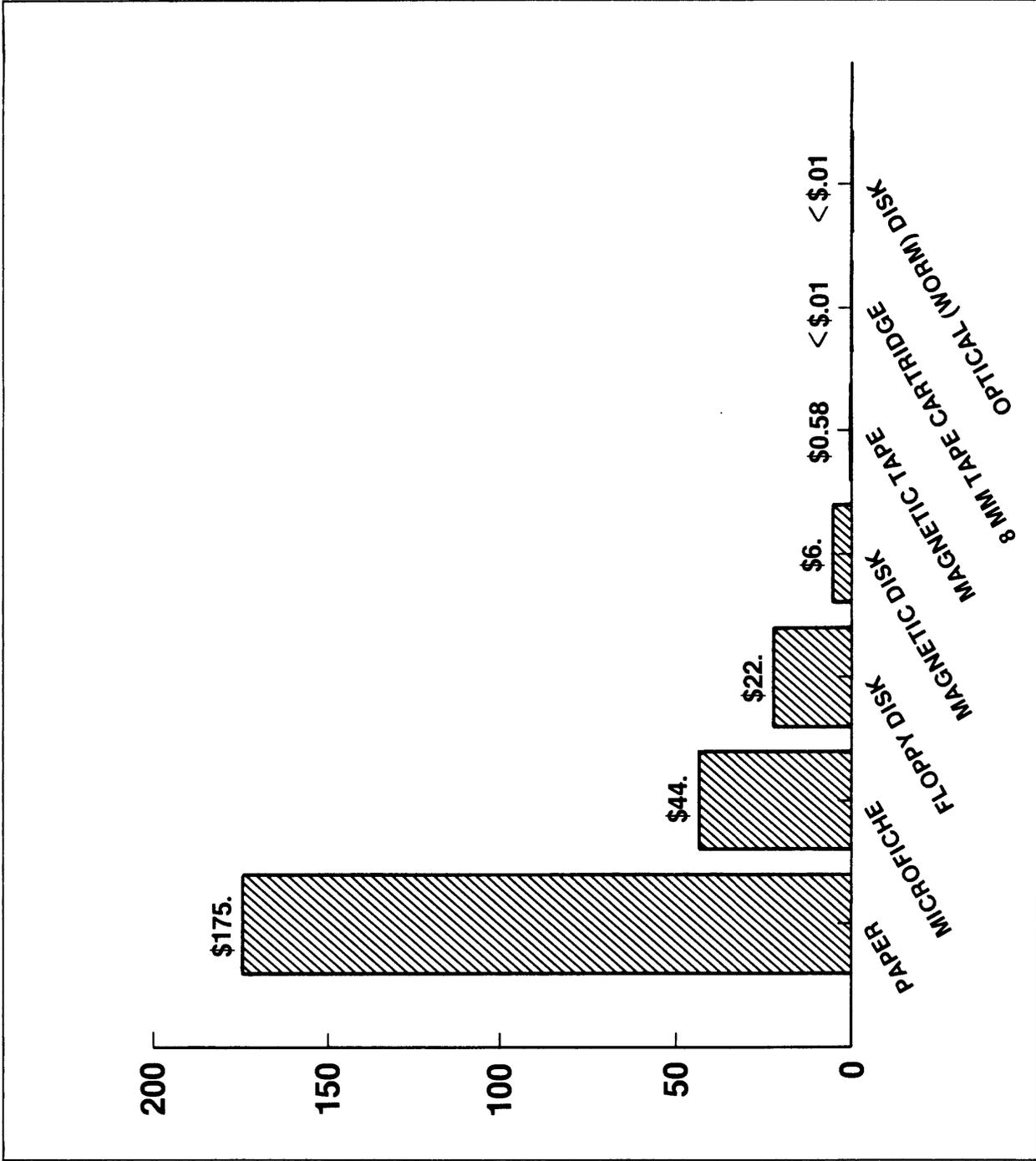


Figure 2. Average annual long-term storage cost for 100 megabytes of data.

- Optical media are less likely to be damaged by improper storage, by stray magnetic fields, or excessive temperatures.
- Data are highly compact. In general an optical device of a given size can store 10 to 50 times more data than a magnetic device of the same size. Therefore, long-term storage costs are substantially less.
- Cost of optical media is less than the cost of magnetic disks or floppy disks.
- Optical disks allow random access of files whereas magnetic tapes allow only sequential access.

Compact Disk Read-Only Memory Compared to Write-Once/Read Many Media

CD ROM 5.25-inch disks hold up to 650 MB on one side. The cost of making the metal master and stamping out 100 CD ROM disks is \$3,000 to \$6,000 (Shier, 1989). Additional disks cost about \$3 each to produce. Plug-in CD ROM readers can be obtained for about \$600. Internal CD ROM readers that fit in a half-height disk drive slot cost about \$500.

The main advantages of CD ROM media are as follows:

- Nearly all disks have the same format and are interchangeable with other manufacturer's disks.
- The cost of the required peripherals is less than for write-once disks.
- Production of large numbers of disks for distribution is inexpensive.

WORM disk drives are able to read and write data. A 5.25-inch WORM drive for a microcomputer ranges in cost from \$3,000 to \$6,000. A 5.25-inch WORM cartridge ranges in capacity from 400 MB to 1.2 GB (gigabytes). A 12-inch WORM

drive for a minicomputer or workstation ranges in cost from \$15,000 to \$20,000. A 12-inch WORM cartridge ranges in capacity from 1.2 to 2.4 GB. The cost and capacity of the disk drive and media vary from manufacturer to manufacturer. WORM drives have the following advantages:

- Most drives are set up so that if a file with an identical name is stored, the previous version is subsequently ignored, unless special software is used to find and copy it (Shier, 1989).
- Generating a new data disk takes much less time. Writing a 5.25-inch WORM disk on a microcomputer by the Microsoft Disk Operating System (MS-DOS) copy command or a simple program requires about 4 hours. The turnaround time to master a new CD ROM takes weeks.
- The cost of optical media and long-term storage of the media are substantially less than any other media except for 8-mm tape cartridge.

One problem with the current WORM media is the lack of a common standard. Although outwardly similar, the optical format and physical dimensions of each manufacturer's products are significantly different from those of its competitors (Bican, 1988). Consequently, one manufacturer's cartridge cannot be used in another manufacturer's drive, or conversely. Standards are evolving slowly and are being addressed first for the media. Another problem is that of legality. Because a document stored on optical disk (CD ROM or WORM) has not been used as evidence in a court of law, agencies are using this as a reason to not "trust" their documents to optical media (Association for Information and Image Management, 1990).

Erasable Compared to Write-Once/Read-Many Media

Erasable media are rewritable and high-capacity disks are available. Erasable optical disks are in development and not yet available in high-volume production quantities. Costs for equivalent erasable optical disk drives and media is estimated to be 20 to 40 percent more than for WORM equipment and media. Because an erasable disk can be altered, it is unacceptable as an archival medium.

Paper and Microfiche Compared to Optical Media

Current U.S. Geological Survey policy calls for the archival of data on paper or microfiche. Paper is the dominate form of information exchange. The advantage of using paper to archive information is that paper is convenient and does not have to be computer encoded before archival. The advantage of microfiche is that this medium is more stable and has a longer life expectancy than paper and other types of media.

The disadvantages of paper and microfiche are:

- Paper requires a relatively large area for long-term storage; microfiche much less.
- The area of long-term storage needs to be climate-controlled.
- When compared to optical media, a substantially large amount of paper or microfiche is required to archive information.
- The cost of the paper and microfiche media and long-term storage is substantially greater than optical media.
- It is much more difficult and time consuming to locate a particular piece of information on paper media or microfiche media compared to optical media.
- Data in digital form are readily available for use on a computer. Information archived on paper or microfiche generally is entered into the computer by hand.

ADVANTAGES OF WRITE-ONCE/READ-MANY OPTICAL STORAGE

WORM technology offers many advantages, some of which are (Storage Dimensions, 1989; and Maximum Storage, Inc., written commun. Dec. 15, 1989):

- Permanence -- Write-once technology offers virtually permanent data storage that cannot be accidentally erased or damaged through static electricity,

magnetic fields, or other environmental factors, which can destroy standard-fixed-magnetic disks or tape storage. The data on the write-once cartridge can never be overwritten or erased, accidentally or otherwise.

- **Longevity** -- Data stored on write-once media has a guaranteed shelf life of at least 10 years, considerably longer than its magnetic counterparts.
- **Durability** -- The actual disks are contained in a sturdy cartridge, providing a secure environment and protecting them from damage. Cartridges can be stacked or placed side-by-side.
- **Capacity** -- Data storage capacities per cartridge range from 400 MB to more than 1 GB in a cartridge measuring 5-1/4" x 6" x 3/8".
- **Removability** -- Write-once technology is the most secure and compact removable data storage available. The permanent nature of write-once storage makes it impervious to many of the natural elements that have affected removable disk and tape storage in the past.
- **Random Access** -- No longer must archive data be stored on sequential access media such as magnetic tape. With the random access of optical disks, the long delays in searching tapes no longer occur.
- **Transportability** -- Because of the small size, high-storage density, and relative ruggedness of a 5.25-inch cartridge, it can be used to distribute large amounts of data conveniently and economically between sites.
- **Cost Effectiveness** -- With media cost now less than \$0.20 per MB, it is a highly-competitive form of data storage.

ARCHIVAL PROCEDURE AND RESULTS

The Automated Data Processing System (ADAPS) Utility Program, UV__ARCHIVE, of the National Water Information System (NWIS), was executed on a Prime 9955II minicomputer to archive all unit-value data for the 1985 water year from the Tampa Subdistrict Office of the USGS. The UV__ARCHIVE program

created over more than 400 unit-value data files. The amount of disk space used was about 106 MB. Each file contains detailed information about the type of data collected, followed by all the unit-value data for one gaging station. The unit-value data files are stored in fixed-column American Standard Code for Information Interchange format (fig. 3).

The first test was conducted on an optical-disk subsystem, Model 525WC, Information Storage, Inc. (ISI), connected to a COMPAQ 286 microcomputer. The optical-disk subsystem emulates a magnetic disk, and uses a 5.25-inch disk with a capacity of 115 MB per side. A Fortran program was written and executed on the Prime to create an index file (fig. 4) of the 400 unit-value data files. Each record in the index file contains general information about an individual unit-value data file including an alternate filename, the original UV__ARCHIVE file name, station name, station number, station type, location, type of unit-value data collected, amount of record collected, latitude, longitude, and so forth.

To store the unit-value data files on an optical disk, the file names had to be modified to correspond to MS-DOS naming conventions. A Prime Command Procedure Language (CPL) program was developed to rename the archived data files on the Prime using the alternate file name in the index file. The archived data files were transferred to the microcomputer magnetic disk by a program called FTP (File Transfer Protocol). This program provides a simple way to transfer files between a local computer and a remote computer. The MS-DOS copy command was used to transfer the archived files from the microcomputer magnetic disk to the optical disk.

The second test was conducted on an optical disk subsystem, Model 8502, Dallastone, Inc., connected to a Prime 6550 minicomputer. The optical disk subsystem emulates a tape drive, and uses a 12-inch optical disk with a capacity of 1 GB per side. The same Fortran program was executed to create the index file. To store the unit-value data files on the optical disk, a CPL program that wrote the unit-value data files to the Dallastone optical disk using the Prime Operating System (PRIMOS) tape utility MAGSAV was executed. The time to transfer the data from the magnetic disk to the optical disk (in both tests) was comparable to writing files to

Code	Agency	Station	Parameter	Value	Code	Agency	Station	Parameter	Value
BE	STDARC								
DB	01								
SD	USGS	274225081315201							
SE	2	7202000011	9	V					
TM	19841001								
UA	5010000	2129	0a	020000	2129	0a	040000	2129	0a
UA	5060000	2129	0a	070000	2129	0a	090000	2129	0a
UA	5110000	2129	0a	120000	2121	0a	140000	2121	0a
UA	5160000	2121	0a	170000	2121	0a	190000	2121	0a
UA	4210000	2121	0a	220000	2121	0a	240000	2129	0a
TM	19841002								
UA	5010000	2129	0a	020000	2129	0a	040000	2129	0a
UA	5060000	2128	0a	070000	2128	0a	090000	2128	0a
UA	5110000	2128	0a	120000	2128	0a	140000	2128	0a
UA	5160000	2128	0a	170000	2120	0a	190000	2128	0a
UA	4210000	2128	0a	220000	2127	0a	240000	2127	0a
TM	19841003								
UA	5010000	2127	0a	020000	2127	0a	040000	2127	0a
UA	5060000	2127	0a	070000	2127	0a	090000	2127	0a
UA	5110000	2127	0a	120000	2127	0a	140000	2127	0a
UA	5160000	2126	0a	170000	2126	0a	190000	2126	0a
UA	4210000	2126	0a	220000	2126	0a	240000	2126	0a
TM	19841004								
UA	5010000	2126	0a	020000	2126	0a	030000	2126	0a
UA	5060000	2126	0a	070000	2126	0a	080000	2126	0a
UA	5110000	2125	0a	120000	2125	0a	130000	2125	0a
UA	5160000	2121	0a	170000	2121	0a	180000	2121	0a
UA	4210000	2125	0a	220000	2125	0a	230000	2125	0a
TM	19841005								
UA	5010000	2124	0a	020000	2120	0a	030000	2120	0a
UA	5060000	2124	0a	070000	2124	0a	080000	2124	0a
UA	5110000	2124	0a	120000	2120	0a	130000	2120	0a
UA	5160000	2124	0a	170000	2123	0a	180000	2123	0a
UA	4210000	2123	0a	220000	2123	0a	230000	2123	0a
TM	19841006								
UA	5010000	2123	0a	020000	2123	0a	030000	2123	0a
UA	5060000	2123	0a	070000	2122	0a	080000	2122	0a
UA	5110000	2122	0a	120000	2122	0a	130000	2122	0a
UA	5160000	2122	0a	170000	2122	0a	180000	2122	0a
UA	4210000	2122	0a	220000	2122	0a	230000	2122	0a
TM	19841007								
UA	5010000	2121	0a	020000	2121	0a	030000	2121	0a
UA	5060000	2121	0a	070000	2121	0a	080000	2121	0a
UA	5110000	2121	0a	120000	2121	0a	130000	2121	0a
UA	5160000	2120	0a	170000	2120	0a	180000	2120	0a
UA	4210000	2120	0a	220000	2120	0a	230000	2120	0a
TM	19841008								
UA	5010000	2120	0a	020000	2120	0a	030000	2120	0a
UA	5060000	2120	0a	070000	2120	0a	080000	2120	0a
UA	5110000	2119	0a	120000	2119	0a	130000	2119	0a
UA	5160000	2119	0a	170000	2119	0a	180000	2119	0a
UA	4210000	2119	0a	220000	2119	0a	230000	2119	0a

Code	Agency	Station	Parameter	Value	Code	Agency	Station	Parameter	Value
BE			Begin Envelope						
DB			STDARC						
			Data Base						
			SD Station Data						
			USGS						
			274225081315201						
			SE Sensor Information						
			2						
			7202000011						
			9						
			V						
			E						
			TM Time Information						
			19841001						
UA			Unit values -- Archive						
			5010000						
			020000						
			030000						
EE			End Envelope						

EXPLANATION

Standard format; unit values archived.
 Data-base identification number.
 Agency; station number.
 Data descriptor number; 72020 - parameter code; 00011 - statistic code; length of data field; variable interval recording; edited unit values.
 Date of first value (YYYYMMDD).
 5 - Number of unit values in record;
 010000 - HHMMSS first value recorded;
 2129 - unit value (specified in SE);
 0 - write protect code;
 a - rounding code.
 second value recorded;
 third value recorded....
 End.

Figure 3. Sample of unit value archival data file.

```

*****
273851082031503 ROMP 40 SH WELL NEAR DUETTE,FL
UDB01.USGS.273851082031503.1985 W1985_01.244
GROUND WATER OTHER THAN SPRING FLORIDA POLK
273851 0820315 03100101 FLORIDA
72020 ELEVATION (FEET NGVD) 122BNVL
00011 RANDOM INSTANTENOUS
365 0
*****
274031082150401 ROMP DP WELL 123
UDB01.USGS.274031082150401.1985 W1985_01.245
GROUND WATER OTHER THAN SPRING FLORIDA HILLSBOROUGH
274031 0821504 03100203 FLORIDA
72020 ELEVATION (FEET NGVD) 52.00
00011 RANDOM INSTANTENOUS
243 122
*****
274208082063301 JOYCE RAINGAGE NEAR FORT LONESOME,FL
UDB01.USGS.274208082063301.1985 W1985_01.246
METEOROLOGICAL FLORIDA HILLSBOROUGH
274208 0820633 03100203
00045 RAINFALL ACCUMULATED (INCHES)
00006 SUM
11 354
*****
274213082043401 HARRISON RAINGAGE NEAR FORT LONESOME,FL
UDB01.USGS.274213082043401.1985 W1985_01.247
METEOROLOGICAL FLORIDA HILLSBOROUGH
274213 0820434 03100203
00045 RAINFALL ACCUMULATED (INCHES)
00006 SUM
11 354
*****
274215082072000 UNNAMED TRIB @ CNTL SITE SRD674 NR FT LONESOME
UDB01.USGS.274215082072000.1985 W1985_01.248
STREAM FLORIDA HILLSBOROUGH
274215 0820720 03100203
00065 GAGE HEIGHT, FEET
00011 RANDOM INSTANTENOUS
357 8
*****
274275081315201 USGS SHALLOW WELL P-48 NEAR FROSTPROOF, FL
UDB01.USGS.274275081315201.1985 W1985_01.249
GROUND WATER OTHER THAN SPRING FLORIDA POLK
274275 0813152 03090101 FLORIDA
72020 ELEVATION (FEET NGVD) 144.31
00011 RANDOM INSTANTENOUS
360 5
*****

```

EXPLANATION	
Station number	station name
Original filename	alternate filename
Site type	state district county
Lat. Long.	hydrologic code geologic code gage datum
Parameter code	parameter name
Statistic code	statistic name
No. days with data	No. days without data

Figure 4. Sample of unit value index file.

a magnetic tape. Access time to retrieve any data file from the 525WC was a matter of seconds; to retrieve file number 400 on the Dallastone 8502 was 7 minutes. The Dallastone is a tape emulation system and cannot access files randomly like magnetic disk emulators.

SUMMARY

In the tests conducted, optical storage proved to be the most efficient and easy means of archiving unit-value data. Three types of optical devices are currently available: compact disk read-only memory (CD ROM), write-once/read-many (WORM), and erasable. The costs associated with archiving information on several different types of media were examined. It was determined that paper was the most expensive medium, followed by magnetic disk, microfiche, floppy disk, WORM, magnetic tape, and 8-mm tape cartridge. The long-term storage costs for paper, microfiche, floppy disk, magnetic disk, and magnetic tape are substantially more than for optical disk or 8-mm tape cartridge. The long-term storage costs of both the optical and 8-mm tapes are comparable, with both being less than \$0.01 per year for 100 MB of storage.

Because write-once disks cannot be erased, they are more suitable for applications involving data archiving and as backup storage devices than magnetic, CD ROM, and erasable optical media. Magnetic media are easily damaged by static electricity, stray magnetic fields, or excessive temperature or humidity. The CD ROM media is designed for the dissemination of information. Information stored on erasable optical media can be easily altered.

Paper and microfiche are alternative media choices that are currently being used to archive data. However, the cost of these media and the cost to maintain them indefinitely is substantially greater than for optical media. One key factor in favor of data stored on optical disk is that these data are readily accessible by computers, whereas data stored on paper or microfiche are accessed manually.

One year of hydrologic data was archived on two different optical-disk subsystems. One was a disk emulation system and the other was a tape emulation system. Both systems performed satisfactorily. Files on the disk emulation system, however, were accessed in seconds, whereas it took 7 minutes to access the 400-hundredth file on the tape emulation system.

Although no media for data storage can be considered as permanent, the high data-storage capacities, long shelf life, nonerasability, and low cost of optical write-once media make it a suitable media for archiving hydrologic data.

REFERENCES

- Association for Information and Image Management, 1990, Yates VS Crates:
A solution to any storage problem: January 1990 Newsletter, National Capitol
Chapter, 16 p.
- Bican, Frank, 1988, WORM Drives: PC Magazine, March 29, 1988.
- Owen, D. J., 1989, ICI Imagedata's Digital Paper: Optical Information Systems,
September-October 1989, v. 9, no. 5, p 226-229.
- Shier, Daniel, 1989, Optical Disk Ideas on the Rise of Exploration: Geobyte,
February 1989, p 6-14.
- Storage Dimensions, 1989, Write-Once Optical Storage for Personal Computers:
A White Paper, 30 p.