United States General Accounting Office

GAO

Fact Sheet for the Chairman, Subcommittee on Science, Technology, and Space, Committee on Commerce, Science, and Transportation, U.S. Senate

September 1990

SPACE DATA

Information on Data Storage Technologies

RESTRICTED——Not to be released outside the General Accounting Office unless specifically approved by the Office of Congressional Relations.

GAO

United States
General Accounting Office
Washington, D.C. 20548

Information Management and
Technology Division

B-240617

September 12, 1990

The Honorable Albert Gore, Jr.
Chairman, Subcommittee on Science,
Technology, and Space
Committee on Commerce, Science,
and Transportation
United States Senate

Dear Mr. Chairman:

As requested by your office, we are providing information on current and advanced data storage technologies to assist the committee in evaluating their potential use for the National Aeronautics and Space Administration's (NASA) future storage needs. Specifically, you requested that we identify the general characteristics and costs of these data storage technologies. In a future report we will furnish information on NASA's plans for using and applying these technologies to store the large amounts of space science data expected from the growing number of missions scheduled for the 1990s.

The National Aeronautics and Space Act of 1958[1] placed responsibility on NASA for conducting space exploration research that contributes to the expansion of human knowledge and directed it to provide the widest practicable and appropriate dissemination of this information. Since 1958 NASA has spent more than $24 billion on space science to help us understand our planet, solar system, and the universe. It has launched over 260 major space science missions and has acquired massive volumes of data. The majority of data from these past missions is stored on at least 1.2 million magnetic tapes, which have the capacity to store over 90 billion pages of text.[2]

NASA anticipates the volume of data generated and stored for future missions will be unparalleled in the history of the agency. It estimates that the annual volume of archived data will rise from 63 terabits[3] in 1990 to more than 4,200 terabits by the late 1990s—more than a 6,500-percent

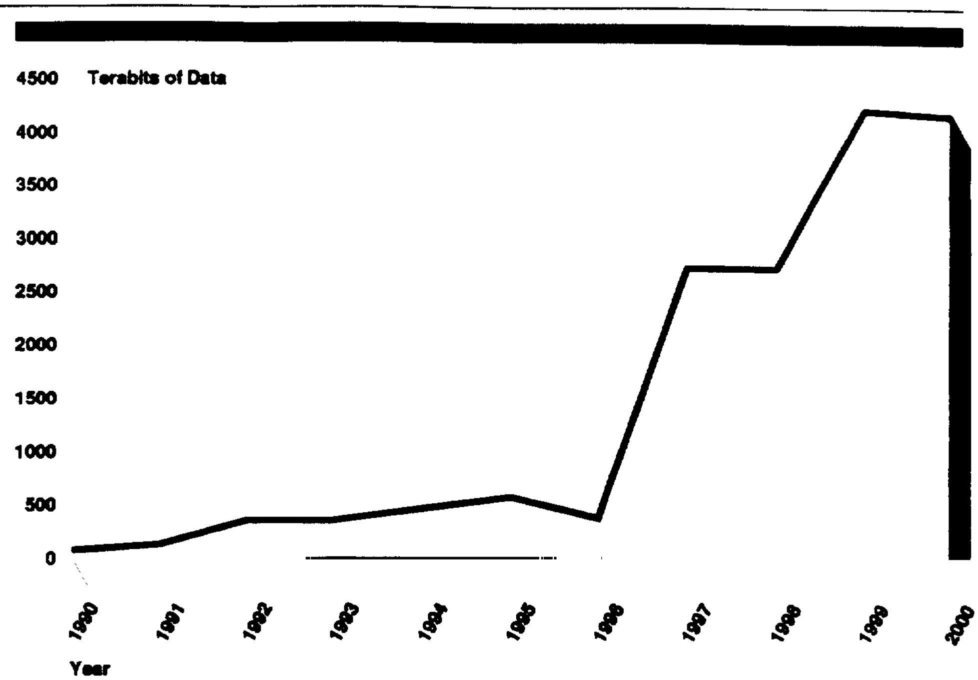
[1] Public Law 85-568.

[2] This estimate is based on the storage capacity of a standard 2,400-foot-long tape, with data stored at 6,250 bits per inch. A single page of text contains about 400 words, with 6 characters per word—or about 19,200 bits of data.

[3] One terabit of data is approximately 10^{12} bits (or 1 trillion bits). About 700 high-density tapes (6,250 bits per inch) would be required to store 1 terabit of data.

jump. Figure 1 shows the expected increase in data volumes through the 1990s.

Figure 1: NASA's Estimated Annual Data Volume



Storing such large volumes of data will require new and more efficient data-storage media. In the past few years large strides have been made in both magnetic tape and optical disk technologies with the commercial introduction of higher capacity tape and optical media. Appendix I briefly explains several commercially available data storage technologies. The choice of a data-storage medium depends on the amount of data to be archived, the storage capacity of the medium, and the cost of the medium and supporting equipment needed to access it. Appendix II presents several general performance and archiving characteristics of data-storage technologies. Appendix III presents a detailed breakdown of costs.

The information in this report was obtained from manufacturers' specifications, available literature, and discussions with NASA officials and other experts. Our work was conducted between December 1989 and

July 1990. We discussed the contents of this report with NASA headquarters, Jet Propulsion Laboratory, and Goddard Space Flight Center officials, and have incorporated their comments where appropriate. Details of our objectives, scope, and methodology appear in appendix IV.

As arranged with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution of it until 30 days from the date of this letter. At that time we will send copies to other appropriate congressional committees; the Administrator, NASA; and other interested parties. We will also make copies available to others upon request.

This work was performed under the direction of Samuel W. Bowlin, Director, Defense and Security Information Systems, who can be reached at (202) 275-4649. Other major contributors are listed in appendix V.

Sincerely yours,

for Michael Brzezinski

Ralph V. Carlone
Assistant Comptroller General

Contents

Abbreviations

CD-ROM	compact disk-read only memory
DAT	digital audio tape
GAO	General Accounting Office
GB	gigabyte
IMTEC	Information Management and Technology Division
MB	megabyte
NASA	National Aeronautics and Space Administration
WORM	write once read many

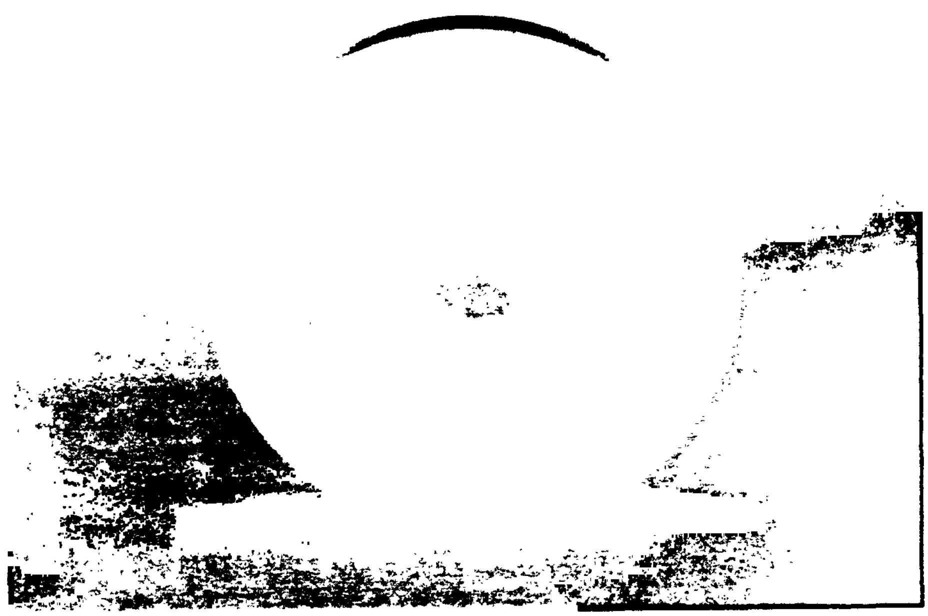
Definitions of Data-Storage Technologies

The following is a brief description of seven data-storage technologies that were commercially available when we did our work. The figures used to portray each technology were provided by several manufacturers and used with their permission. Their use, however, does not imply an approval or endorsement of the manufacturers or their products by us.

Magnetic Reel

One-half-inch wide, reusable, 9-track magnetic tape. Data are written horizontally on the tape and wound on reels. Tape is considered a sequential-access medium, i.e., data records are accessed one after the other, in the order in which they are physically stored on the tape.

Figure I.1: Magnetic Reel

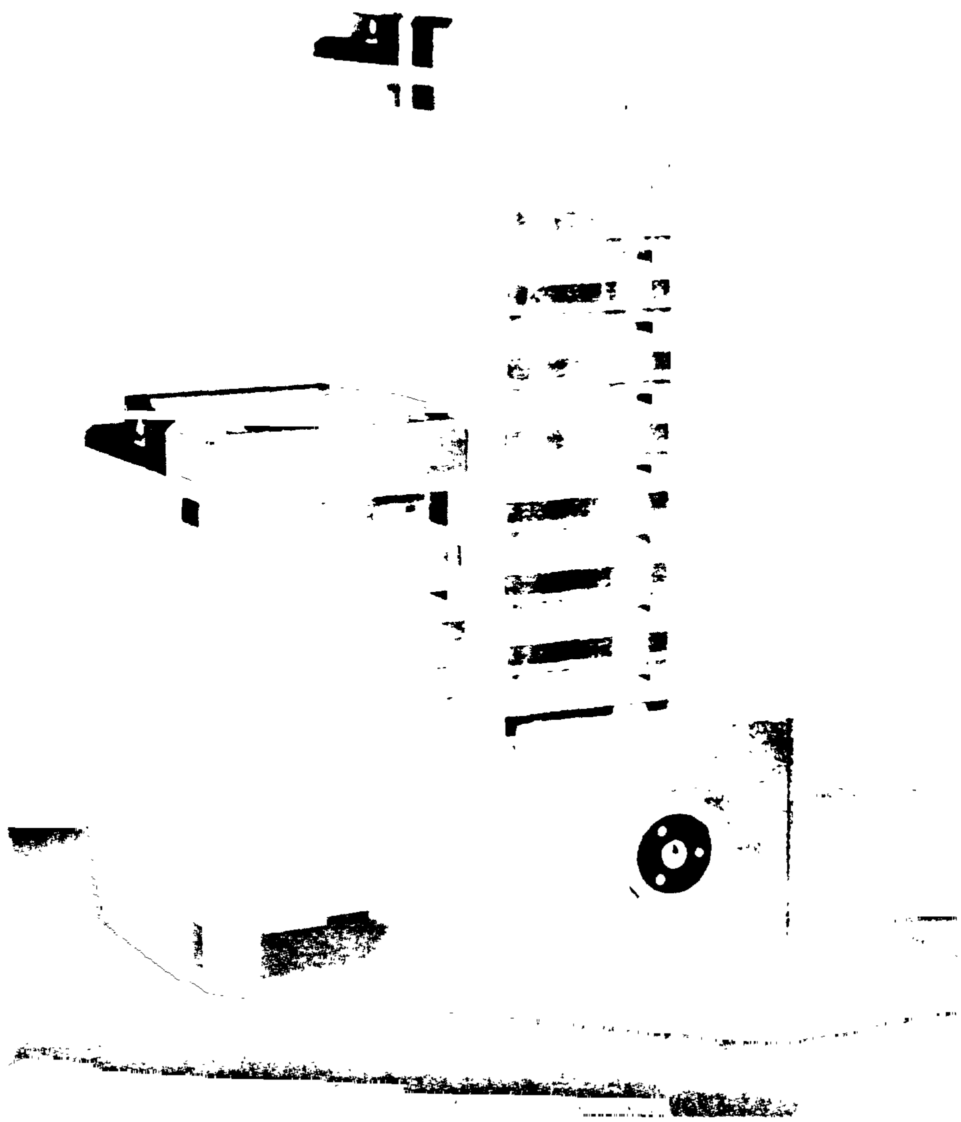


Source: Ampex Corporation.

3480-Compatible

One-half-inch wide, reusable, 18-track magnetic tape, stored in cartridges. Data are written horizontally on the tape. This is a sequential-access medium.

Figure I.2: 3480-Compatible

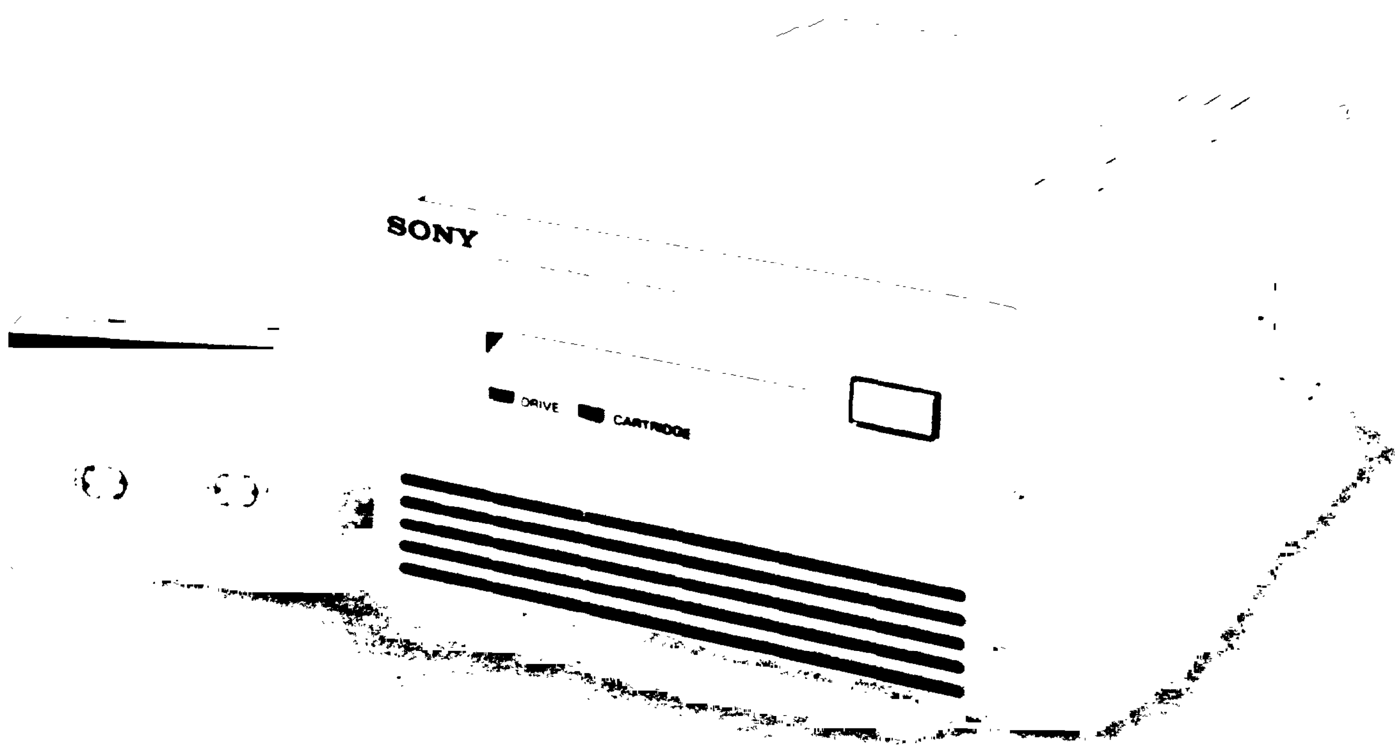


Source: Fujitsu America, Inc.

Helical Scan

This technology uses a rotating head to write data diagonally on reusable magnetic tape. The most common tape sizes are 8 millimeter (mm) tape and 4mm digital audio tape (DAT). Helical scan is a sequential-access medium.

Figure I.3: Helical Scan, 4mm

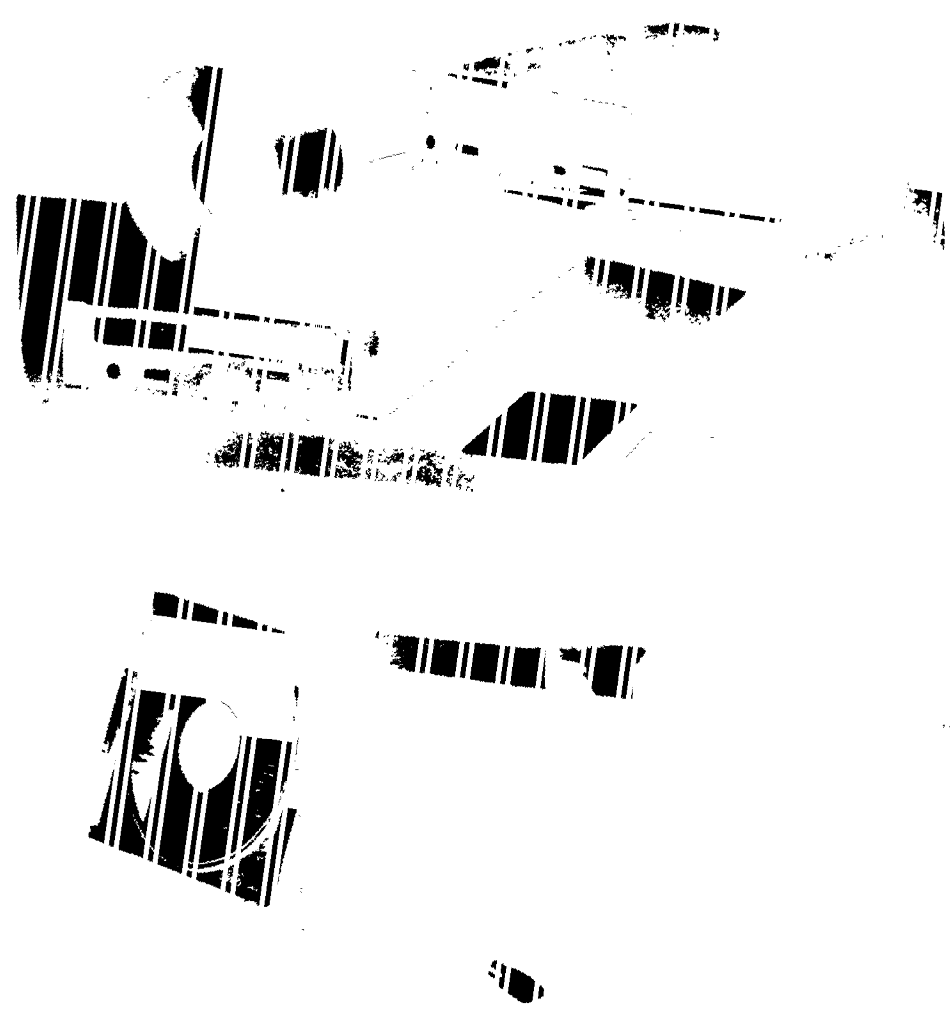


Source: Sony Corporation of America.

Compact Disk-Read Only Memory (CD-ROM)

An optical disk similar to the CD audio disk. Data are written by laser on master disks, which are then typically produced and duplicated commercially. Compact disks can only be written once and data can be accessed randomly (data records are written or accessed directly without sequentially searching through other data records first).

Figure I.6: Compact Disk-Read Only Memory

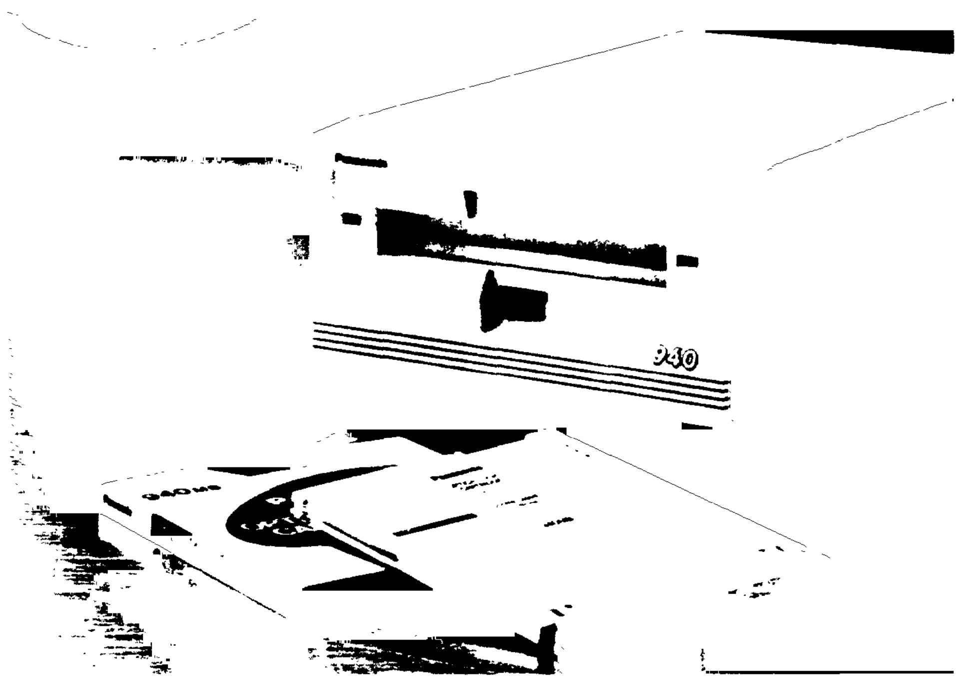


Source: Sony Corporation of America.

Write Once Read Many (WORM)

The WORM is an optical disk on which data are written by laser. The disk can only be written on once and is accessed randomly.

Figure I.5: Write Once Read Many



Source: Panasonic Communications and Systems Company, Division of Matsushita Electric Corporation of America.

Rewritable Optical Disk

A magneto-optical disk that combines magnetic and optical (laser) recording technologies. The disk can be written on many times and is accessed randomly.

Figure I.6: Rewritable Optical Disk

Source: Sony Corporation of America.

Optical Tape

A product made from digital paper, a thin flexible film manufactured in large rolls and cut into many shapes and sizes. Data are written once by laser on a nonreusable tape. This is a sequential-access medium.

Figure I.7: Optical Tape

Source: ICI Imagedata.

General Characteristics[a]

	Technology	
Characteristic	**Magnetic reel**	**3480-compatible[b]**
Storage capacity	**Low**	**Low**
	.09-.2GB	.2GB
Access time	**High**	**Med**
	1-5 min	13-25 sec
Transfer rate (per second)	**Low/Med**	**Med/High**
	.3-1.25MB	1.5-4.5MB
Storage life[c]	**Low**	**Low**
	3-10 yrs	3-10 yrs
Bit error rate	**High**	**Low/High**
	10^{-10}	10^{-13}; 10^{-11}
Format standards?	**Yes**	**Yes**
Mature technology?	**Yes**	**Yes**

Technology							
Helical 4mm (DAT)	**Helical 8mm**	**CD-ROM**	**WORM 5.25"**	**WORM 12"**	**WORM 14"**	**Rewritable optical disk**	**Optical tape**
High	**High**	**Med**	**Med**	**High**	**High**	**Med**	**High**
1.3GB	2-5GB	.5-.8GB	.6-.9GB	2-6GB	7-8GB	.6-1GB	1000GB
Med	**High**	**Low**	**Low**	**Low**	**Low**	**Low**	**Med**
20 sec	10 min	150-800 ms	60-250 ms	145-250 ms	9-700 ms	35-95 ms	28 sec
Low	**Low**	**Low**	**Low/Med**	**Low**	**Med**	**Low/Med**	**High**
.2MB	.5MB	.15MB	.3-1.3MB	.3-.8MB	1MB	15-1.5MB	3MB
Low	**Low**	**Med/High**	**Med**	**Med**	**Med**	**Med**	**Med**
3-10 yrs	3-10 yrs	20-100 yrs	10-30 yrs	10-30 yrs	10-30 yrs	10+ yrs	15+ yrs
Low	**Low**	**Med**	**Med**	**Med**	**Med**	**Med**	**Med**
10^{-15}	10^{-13}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
No	**No**	**Yes**	**No**	**No**	**No**	**No**	**No**
No	**No**	**Yes**	**No**	**No**	**No**	**No**	**No**

[a]Manufacturer specifications were used as the information source in most cases. The actual, effective values may be different. For example, the bit error rate does not include errors caused by hardware or handling problems. Transfer rates have been reported by some users to be as low as half of what manufacturers claim.

[b]There are two versions of 3480 drives.

[c]Optical technologies have not been commercially available for data storage long enough to know if they will meet the claims for storage life.

Definitions

Storage capacity: the amount of data that can be stored on the medium, measured in gigabytes (GB). A gigabyte is approximately one billion bytes of data, and one byte is 8 data bits. For archiving, higher capacity is important for storing a high volume of data.
Access time: the interval between a request to read or store data and the completion of that task. Time is measured in minutes (min), seconds (sec), or milliseconds (ms). A millisecond is one thousandth of a second. For archiving, access time is less important than other characteristics.
Transfer rate: the rate at which data are transferred from the drive to the computer, measured in megabytes (MB) per second. A megabyte is one million bytes of data. For archiving, a high transfer rate is important when frequent requests or large volumes of data must be accessed from storage.
Storage life: the period of time data will remain usable given reasonable care and maintenance of the media. For archiving, media with a long storage life reduces the frequency of recopying the data over time.
Bit error rate: the probability of a bit of data delivered from the device being incorrect. For archiving, it is important that data that will be stored for future reference and analysis be correct and therefore have a low bit error rate.
Format standards: rules describing how the data are stored. For archiving, format standards are especially important because the data will be used over a long period of time, probably by many users.
Mature technology: a technology that is readily available on the commercial market and that has been in operational use in many installations over a substantial period of time. For archiving, a mature medium is important to ensure that it can be kept and read easily and accurately over a long period of time.

Costs of the Technologies[a]

Technology	Cost per megabyte[b]	Media unit cost[c]	Drive cost
Magnetic reel	$0.11–.16	$14–21	$5,000–20,000
3480-compatible[d]	.03–.08	6–16	15,000–90,000
Helical 4mm	.02–.03	20–35	4,000–7,000
Helical 8mm	.003–.008	6–40	3,000–7,000
CD-ROM	.02–.20	1500 master/ 2 copy	445–3,695
WORM 5.25"	.11–.33	65–298	2,500–10,000
WORM 12"	.07–.17	345–400	13,000–17,000
WORM 14"	.07–.09	595	25,000
Rewritable optical disk	.42–.65	249–650	4,500–10,950
Optical tape	.01	10,000	200,000

[a]These figures are for comparison purposes only. We limited our review to the media and drive costs. Other hardware and software to access and use the storage medium would be needed, depending on the technology selected. For example, one large mass storage system we identified included a supercomputer, numerous user workstations, an operations workstation, magnetic disks, printers, a local area network, and a smaller computer known as a file server, which essentially permits components of the system to access the mass storage system.

[b]The cost per megabyte was calculated by dividing the minimum cost for the storage technology (disk, tape, etc.) by the minimum capacity for the low end of the range, and dividing the maximum cost and capacity for the high end of the range.

[c]This represents the cost for one unit of storage media, i.e., one tape reel, cartridge, or disk

[d]There are two classes of 3480 drives: large, and the smaller "rack mountable" version. The low-end pricing is for the smaller version, and high end for the larger version.

Objectives, Scope, and Methodology

The Subcommittee on Science, Technology, and Space, Senate Committee on Commerce, Science, and Transportation, asked us to identify (1) relevant advanced data storage technologies, including their strengths, weaknesses, and costs; (2) current NASA initiatives in this area; and (3) major agency programs that may benefit from the use of these technologies. As agreed in a subsequent meeting with the Subcommittee, this report provides information on the first area, relevant technologies and their attributes, in summary form; information on the second and third areas will be presented in a follow-up report.

Information on the estimated volumes of data to be generated through the year 2000 was obtained from the NASA Office of Space Science and Applications.

To identify and describe applicable data storage technologies, we

- conducted a literature search of articles describing advanced data archiving technologies;
- contacted manufacturers of advanced data archiving technologies to obtain product background and specifications;
- reviewed NASA studies, reports, and other documents related to data management prepared by various scientific groups and committees;
- interviewed NASA, Jet Propulsion Laboratory, and Goddard Space Flight Center officials (including National Space Science Data Center officials) responsible for managing and overseeing NASA's data;
- conferred with data storage media experts; and
- attended a symposium on mass storage systems sponsored by the Institute of Electrical and Electronics Engineers.

We reviewed articles in the literature and studies as a basis for identifying relevant data-storage technologies. We limited inclusion to those technologies that were available commercially and capable of handling large volumes of data.

Characteristics and cost information on each technology was gathered from all sources identified above, and inconsistencies in reported attributes were resolved. Manufacturers' specifications were used as a primary data source. Information on technology performance was accepted as provided; no testing of equipment was done to verify the accuracy of manufacturers' claims or other reports of performance.

To compare the costs of the various technologies we obtained price information on the media and the drive required to access the data. An actual storage system requires much more than these components; however, the wide variety of ways in which a complete storage system could be assembled limited us from providing cost information on complete systems.

Experts in data archiving, data processing, magnetic media, and space data storage reviewed our technology definitions and attributes for their accuracy and appropriateness.

Our audit work was performed in accordance with generally accepted government auditing standards, between December 1989 and July 1990, at various locations including NASA headquarters in Washington, D.C.; the Goddard Space Flight Center in Greenbelt, Maryland; and the Jet Propulsion Laboratory in Pasadena, California.

Major Contributors to This Report

Information Management and Technology Division, Washington, D.C.

Ronald W. Beers, Assistant Director
Yvette Ramos, Staff Evaluator

Los Angeles Regional Office

Allan Roberts, Regional Management Representative
George Vindigni, Evaluator-in-Charge
Monica L. Kelly, Staff Evaluator
James D. Nolan, Staff Evaluator
Shawnalynn R. Smith, Staff Evaluator

Ordering Information

The first five copies of each GAO report are free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20877

Orders may also be placed by calling (202) 275-6241.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100