TRANSITIONAL LICENSING SUPPORT SYSTEM

Comparison of Micrographics and Optical Disk Storage and Retrieval Technologies

Prepared for:

NUCLEAR REGULATORY COMMISSION
Division of Waste Management

January 8, 1988

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SECTION 1.0 INTRODUCTION

The purpose of this paper is to examine and compare optical disks and microform media as methods of creating an archive of document images, in order to determine the preferred method. If no clear preference emerges, it is desired to identify conditions and circumstances where one method may be preferred over the other.

This paper presents the advantages and disadvantages of the two media, and endeavors to establish areas of preference where optical disks may have a significant edge over microform and vice versa. The question of optical disks versus microform is timely. Optical disks are a comparatively recent entry into the marketplace, whereas microfilm and microfiche have been a medium of choice for many years. In many archival applications, the question arises of whether to continue to use microform media or to change to digital optical disks.

This study makes a comparison based both on earlier reports, and the latest information in optical disk and microform storage and retrieval, in order to identify the conditions under which each method might be preferred.

The discussion is presented in four sections:

- A description of the archival methods enumerating the functions and discussing the main processes in each step.

- A comparison identifying the advantages and disadvantages of the two methods

- A derivation of the costs of each method

- The conclusion identifying the preferred areas of use for each method

SECTION 2.0 DESCRIPTION OF METHODS

This section presents a step-by-step description of the two archival methods.

2.1 OPTICAL DISKS

2.1.1 Overview

Several variations of optical disk archive systems are used; however, this discussion focuses on an image archive and full-text retrieval system. Such systems are currently installed at the Nuclear Regulatory Commission, the National Agricultural Library, and certain military organizations, and are gaining in acceptance among those who have seen them perform.

The image archive and full-text retrieval system creates an archive of document images, converts text from the images into machine readable form, performs full-text indexing of the text, and enables retrieval of the stored image. The system scans documents and makes an archive file of the images. It then performs recognition processing to convert the text in the image into machine-readable form. Full-text indexing is done and the resulting file is linked to the original image file. The main functions and components of the system are shown in the flow diagram in figure 1.

2.1.2 Scanning

Input to the archive system is hardcopy of diverse types - text, drawings, diagrams, pictures, photos. The first step is

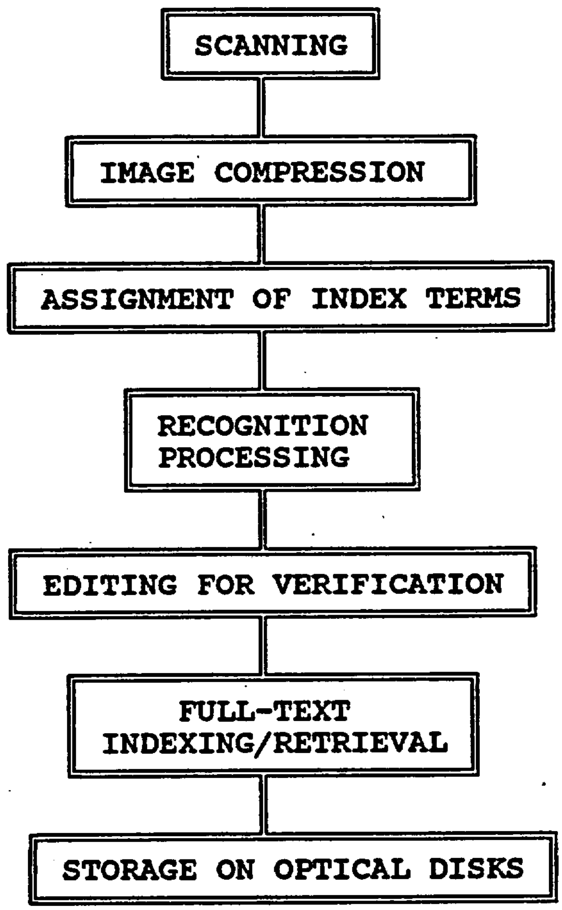

Figure 1 - Flow Diagram of Image Archive and Full-text Retrieval System

scanning, which creates a digital bit-mapped image of the source document. Scanning resolution is 300 dots per inch. Document throughput on the scanner is normally approximately 8 pages per minute.

2.1.3 Image Compression

The uncompressed image of an 8-1/2" x 11" page occupies one megabyte of storage. Compression of the image using CCITT Group 4 algorithms reduces the amount of storage needed to 50 to 100 kilobytes, depending on how "busy" the original document is.

2.1.4 Assignment of Index Terms

A relational database is used to store index terms and accession numbers that organize and categorize the database. Any scheme of index terms can be used that suits the retrieval requirements of the users. Typically, title, author, date, type of document, subject, and sometimes an abstract are included as indices in the relational database. The assignment of index terms and entry of an abstract are manual keyboard processes. The system also automatically generates and assigns an accession number.

2.1.5 Recognition Processing

The bit-mapped image is transferred to a "recognition engine", which converts the bit-mapped image into machine-readable text in ASCII format. This process takes the array of dots in the bit-mapped image and recognizes patterns that form letters and numbers over a broad range of type fonts and sizes. The output of this process is an ASCII data set of the text that entered the process as a bit-mapped image. As an ASCII data set, the text can be subsequently processed by machine. It can be edited using word processors and it can be indexed and searched using commercially available full-text search software.

The processing rate for a recognition engine is at present approximately 50 - 60 seconds per page. New equipment to be released in early 1988 will greatly speed up the processing and increase the rate from one page per minute to 3 or more pages per minute.

2.1.6 Editing for Verification

Although the recognition processor operates with a high degree of success, it is not perfect. The manufacturer claims a successful conversion rate of 99.9% for clean, crisp copy. If the characters are fuzzy or touching or if the background is smudgy or speckled, the error rate increases. Some amount of manual editing is needed to correct the errors so that any further processing can be done on an error-free base. Error correction typically takes approximately three minutes per page.

2.1.7 Full-text Indexing/Retrieval

Full-text indexing and retrieval are complementary processes accomplished through the same commercially available software package. The full-text software builds an inverted file identifying the location of each occurrence of every word (except designated reject words that would never form the basis of a search).

Retrievals are made using the inverted file to locate the pages containing the desired text. Searches can be based on a single word, logical combinations of words, or proximity of words. Partial words ("wild cards") may also be used.

Full-text indexing is done automatically in the batch mode at a cost that is only the cost of running the machine.

A special feature of the image archive and full-text retrieval system is that it can link the ASCII text with the bit-mapped image of the same page. This provides an advantage over other systems because it enables users to search for and locate a page based on a full-text search of the machine-readable code; then they can retrieve the image of that page with a single keystroke. Thus users can see the actual image of a page as it entered the system and view any drawings, diagrams, pictures, formulas, or other matter that could not be converted to machine-readable form.

2.1.8 Storage on Optical disks

Optical disks represent the latest technology in storage media. They are written on by a high energy laser and later read by a low energy laser. They currently operate in a Write-Once-Read-Many-times mode, although manufacturers are preparing to announce erasable media in the first half of 1988. The disks are high volume, high density storage media. The feature of non-erasability makes them suitable for permanent archiving in situations where it is desired to protect and preserve the original version of the information.

Storage volumes are presented in table 1, which presents data for images only, text plus full-text index only, and image plus text plus full-text index.

Table 1 - Capacity of Currently Available Optical Disk Media from Representative Manufacturers

Component	Storage Capacity	Images Only	Text Plus Full-text Index	Image Plus Text Plus Full-text Index
5-1/4" WORM	800 MB	8,000 pp	133,000 pp	7,500 pp
12" WORM	3.2 GB	32,000 pp	533,000 pp	30,000 pp

In table 1, it is assumed that scanning will be performed at a density of 300 dots per inch in each direction, which will result in a bit-map of 8 million bits (1 megabyte) for an 8-1/2 x 11 inch page. Compression processing will reduce the storage needed to an assumed 100 kilobytes. It is also assumed that a page contains 3,000 characters of typed or printed matter that would be converted to ASCII code. The inverted file generated in full-text indexing is assumed to be the same size as the original text, and thus require 3,000 bytes of storage.

Capacity of the optical archive can be extended beyond a single disk by linking multiple single-disk drives or by using a "jukebox".

2.2 MICROFORM DATA STORAGE AND RETRIEVAL

2.2.1 Overview

Microform is the conversion of hardcopy data, principally paper, to film media. Microform technology allows a large volume of source documents to be stored in a compact, stable medium (e.g. silver halide microfilm). There are many types of microfilm, formats, reduction ratios and methods. The equipment required to produce, read, and print images from microfilm also varies.

While film formats and camera types vary in microform, a standard system consists of: microfilm camera and microfilm, an indexing scheme, film image storage, a retrieval scheme, a scanning device, and a reading or reading/printing device. For systems with large amounts of data and/or ones with multiple users a computer-assisted retrieval (CAR) system is needed. This device consists of an electronic microfilm retrieval device and an automatic scanner, both controlled by a microprocessor.

The five components of a microform data storage and retrieval system are shown in figure 2. They are: source document filming, indexing system, image storage, image retrieval, and reader/printer.

Microform systems are designed for a wide range of applications. A minimal system might consist of microfiche, filmed and developed by an independent service bureau. The microfiche is stored in a binder or box. A single

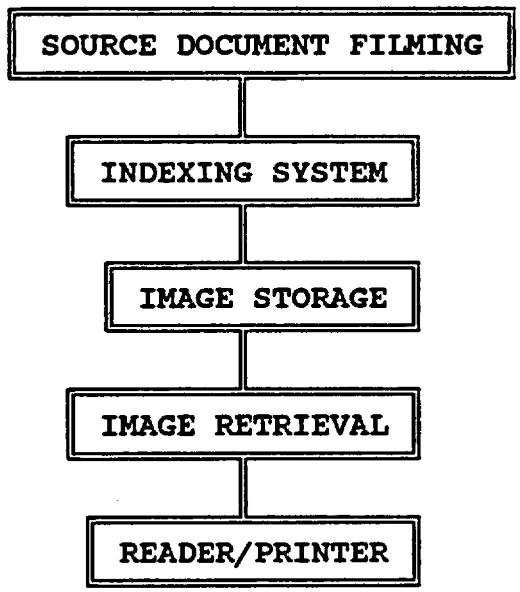

Figure 2 - Microform Storage and Retrieval System

alphabetical index organizes and guides the retrieval. Access and retrieval are accomplished by manually selecting the microfiche, placing it on the carriage of a microfiche reader, and manually moving the carriage under the viewing lens to the desired frame.

At the high end of system capability is a system which is completely automated. The hardware configuration for this system centers around a 32 bit minicomputer which directs and controls all aspects of data storage, retrieval, printing and remote transmission of film images. Users never touch the film and do all their work from a CRT connected to the computer. Filming and developing are done on-site with a high quality camera and processor. The microfilm is stored in cartridges in a large carousel. Multiple indexing is stored in a database in the minicomputer. Access and retrieval are accomplished as follows. The operator queries the database which returns a "hit list" of frame I.D.'s. The operator selects from this list on his screen which causes a robotic mechanism to find the correct microfilm cartridge and load it onto an automatic scanner which is driven by the computer to the desired frame. This image may be hard-copied or digitally scanned, then sent by modem over phone lines to another terminal in the building or to a remote location. Such a system is available at costs approaching $1,000,000.

In between these 2 extremes are microform systems which satisfy the majority of users' needs. This paper will focus on a system which incorporates a microcomputer to store a database index, but one which omits the robotic film loading and remote transmission capabilities. A system of this scale is comparable to the optical disk storage and retrieval system outlined in section 2.1.

2.2.2 Source Document Filming

Microfilm comes in three types: vesicular, diazo, and silver. Silver produces the best quality, but is fragile. Typically, silver film is used to create a master original from which diazo copies are made as more durable working copies.

Prior to filming, the source documents must be prepared. The preparation is determined by the camera type that will do the imaging. In general, extraneous materials are removed, leaving only that which will be filmed.

There are four major camera types that generate the film described above: rotary, continuous form, planetary, and step and repeat. The rotary (flow) camera is a microfilm camera that photographs documents while the document images are being moved by some form of transport mechanism. The document transport mechanism is connected to a film transport mechanism so that the film also moves during exposure and there is no relative movement between the film and the document. The continuous form (rotoline) camera is similar to the rotary camera but is capable of filming continuous forms such as computer printouts. A planetary (flatbed) camera is a microfilm camera in which the document being photographed and the film remain in a stationary position during film exposure. The document is on a plane surface at the time of filming. A step-and-repeat camera is a microfilm camera that can expose a series of separate images on an area of film according to a predetermined format, usually in a grid pattern of rows and columns.

2.2.3 Indexing System

There are five categories of indexing systems: optically coded, computer-controlled page counting, notch and tab, scroll type, and computer-directed frame retrieval.

Optically coded CAR systems encode black-and-white blocks under or between film images. These blocks answer a specific set of questions for every document in the file. The codes are optically searched as the film is driven past the search mechanism. When a match is found the film is stopped on that frame.

In computer-controlled page counting a unique blip (a countable mark) is placed under each frame either during or after filming. A record of each image's blip and frame location is put in the computer index. In retrieval, reference to a frame's blip location enables the reader to automatically position the roll on the desired frame.

Notch and tab indexing is used for unitized storage,i.e., microfiche or aperture card. In this method the card is physically notched at the top in a predetermined order to code filing information. Retrieval can be automatic or manual since the film is ordered by these notches.

In scroll type retrieval a roll (or rolls) of film is stored inside the reader along with an electronically stored index. The computer searches this index and drives the film to the correct frame.

Computer-directed frame retrieval is similar to computer-controlled page counting but no blips are used. Instead the relative position of each frame is recorded in the index.

2.2.4 Image Storage

There are several microfilm formats available. The most common format is roll film in either an open reel or cartridge. The images appear serially along a standard 100 foot long roll. The standard reduction ratio is 24:1, but 100:1 is achievable. 16mm film is typically used for loose documents; 35mm film is used for oversize documents and documents requiring exceptionally high-resolution.

Another format is unitized film. There are two kinds of unitized film: aperture cards and microfiche. Aperture cards are used for single frames, mounted together on a board. The board is keypunched and/or titled for quick reference. This method is usually used for large drawings and plans. Microfiche is a transparent sheet of film with microimages arranged in a 2-dimensional array. A header which can be read without magnification is usually placed in a reserved space at the top edge for indexing and reference. Ultrafiche is a highly-reduced form of microfiche which requires two reduction steps to achieve a 240:1 reduction ratio. Storage capacities for these microforms are summarized table 2.

Table 2 - Capacity of Microform Media

MEDIUM	PAGE IMAGES
Microfiche 24:1	98
Ultrafiche 240:1	1,000
100' Roll Microfilm 24:1	5,000
200' Roll Microfilm 100:1	40,000

2.2.5 Image Retrieval

In the last decade a computer assisted retrieval (CAR) system has replaced the traditional separately developed and maintained microfilm indexing system. This microprocessor-controlled CAR system stores an index in the computer's random access memory (RAM). A CRT terminal is used to query the index. The user then loads the microfilm roll/cartridge into the microfilm reader where an automatic scanner, driven by the microprocessor, drives the film to the correct frame location.

2.2.6 Reader/Printer

Microfilm readers are relatively simple devices whose function is to hold the film in place while a light is projected through the film. A reader/printer combines the reader with an enlarging printer. Microfilm readers are designed for a specific microform corresponding to the microfilm formats identified above. Some, however, are multipurpose. As discussed earlier, additional automating devices, i.e., digitizer, printer, and auto-scanner, may be built into the microfilm reader.

SECTION 3.0 COMPARISON

This section compares the features of the digital system and the microform system on the basis of the following criteria:

- Input rate
- Entry of index terms and abstracts
- Recognition processing and editing
- Indexing/retrieval
- Image quality
- Interfacing to digital systems
- Retrieval time
- Complexity of operation
- Maintenance and cost of operations

3.1 INPUT RATE

The digital system can scan pages at a sustained average rate of 300 pages per hour. Microform systems typically have input rates of 1500 pages per hour.

3.2 ENTRY OF INDEX TERMS AND ABSTRACTS

In either system, the entry of the bibliographic header and index terms is a manual operation, as is the preparation and entry of the abstract. Thus for a given indexing/abstracting scheme there is no difference between the systems. However, because the microform system cannot accommodate full-text search, a more thorough index/abstract scheme may tend to be used for the microform system.

3.3 RECOGNITION PROCESSING AND EDITING

This feature applies only to the digital system. It uses a bit-mapped image as input to the recognition process. The recognition process is done in a batch mode, usually overnight. The day's scanning, which does require an operator to handle the mechanics of the document processing and to oversee system operation, can be recognized and converted to ASCII in unattended operation overnight at low cost.

Editing, however, can become a significant cost item if the original document type is fuzzy, smudgy, or difficult to recognize. Improvements are continually being made in the recognition processors, and one can reasonably expect the error rate to drop and along with it the costs for editing.

3.4 INDEXING/RETRIEVAL

Using index terms and abstracts, searchers can find documents that fulfill the search criteria. To arrive at specific text that they seek they must browse through the entire document. For documents that contain a small number of pages, browsing is not a problem. For longer documents, however, browsing can be burdensome; a search that produces 5 hits of documents that averaged 20 pages each, for example, would require browsing as much as 100 pages to find the material of interest. Both the optical disk system without full-text indexing and the microform system would be limited to this level of retrieval.

Full-text search, which is possible only in the optical disk system, provides the way to get directly to the pages of interest. Browsing can be greatly reduced, and in the particular image archive and full-text retrieval system described in section 2, the images of the selected pages can be brought to the screen immediately, so that searchers seeking drawings, diagrams, pictures, or other graphic material can quickly determine whether they want the page at hand.

In a sentence, microform systems provide document retrieval; digital systems offer page retrieval.

3.5 IMAGE QUALITY

Digital systems provide high quality reproductions. The digital system output device is a laser printer, which prints 300 dots per inch. It produces an output image virtually indistinguishable from the original image; thus little quality is lost in going entirely through the digital system.

Microform media can provide a wide range of quality, which varies as the cost of the system. At the low end are systems like those widely used in libraries that produce readily noticeable degradation in the reproduction. At the high end are systems where the image capture camera resolves to 150 lines per millimeter.

3.6 INTERFACING TO DIGITAL SYSTEMS

The optical disk system already has its information in digital form. It is in the form needed for electronic transmission to a remote location either over phone circuits or local area networks.

Microform systems store their information as pictures. The images must be digitized in order to be transmitted electronically. Some microform systems include digitizers to provide a means of transmitting the selected images at the time they are retrieved.

While both systems compared are expandable, costs rise substantially to do so. Multi-user, remote-access microfilm systems are available at great cost. Multi-user, remote-access optical systems are expected soon, and, while their cost will be higher than single-user systems, for many applications this cost can be offset. Since computers will soon be using optical disks for general purpose mass storage, the optical disk imaging system should be able to migrate into a larger, general purpose, multi-user computer system within a reasonable developmental time frame. All microfilm equipment vendors are actively researching and developing optical storage/retrieval systems for this very reason. The technology is emerging and the future is clearly in this new area. Mechanical microform systems are being replaced by electronic digital systems.

There is an additional factor in this idea of migration to general purpose computer architecture. When digitized, source documents are stored as ASCII, machine-readable information. This means documents processed this way can form input files to general purpose computer programs written in any language supported by the host computer. Text-editors, database, spreadsheet, business, and scientific software will have a new mechanism for inputting data for computation.

3.7 RETRIEVAL TIME

Several differences in retrieval time are noted here. While a microfilm auto-scanner can retrieve and display the desired frame of a mounted cartridge in under 30 seconds, retrieval from a mounted optical disk takes 10 seconds or less. This may impact productivity.

There is a second difference in retrievability. As has been stated earlier, with full-text editing an optical retrieval system will display the precise page containing the desired data --as opposed to the document containing the desired data. This feature would require the additional cost of the optical character recognition hardware, but its benefits are unmatched by any microform system, including the fully-automated one described in section 2 of this report. The ability to perform page level versus document-level retrievals has a significant impact on both total retrieval time and cost. To find data in a large document is time intensive. Also, the cost of printing an entire document can be avoided with page-level retrieval.

Thirdly, there is a great difference in storage capacity between optical disks and microfilm. In our comparison systems a 5-1/4" optical disk versus a 100 foot 24:1 reduction microfilm are configured. In terms of pages of information stored, the optical storage medium has an 8:5 advantage. When full-text indexing is added, however, and only text (as opposed to graphs, pictures, etc.) is stored, there is a 27:1 ratio in favor of optical. All this translates to a real-time retrieval time advantage for optical storage. Any time a retrieval requires access to multiple volumes of storage media, the microfilm system operator will have to change film cartridges 2 to 27 times more often than the optical system operator. This will have a pronounced impact on productivity.

3.8 COMPLEXITY OF OPERATION

In the optical disk storage and retrieval system described in section 2, the workstation is centralized and compact. All steps in the operation, from source document scanning to image retrieval, are immediately accessible. This is not the case with the microform system.

A thriving independent service bureau has grown up around the microform industry. These companies film source documents, process the film, and put it into microfiche jackets or microfilm cartridges for final storage. Why don't customers do their own filming and developing? Cost is an important factor. To achieve high quality film a high quality camera must be used. To achieve professional standards of resolution (150-180 lines per mm) requires a suitable camera. Planetary cameras are able to produce better quality than rotary cameras. They cost from $3,500 - $20,000. Filming is a slow, meticulous activity and labor must be added to this cost.

Next, developing/processing is as important as filming in the production of high quality output A processor costs from $4,000 - $25,000. Again the work is exacting and the labor cost must be added. Also, this step requires use of volatile chemicals. Most sites do not have the proper facilities (or the desire) to store and handle these materials. Last, production control equipment (microscope and densitometer) should be used to verify consistency and accuracy of the results. This equipment typically costs from $1,500 - $5,000. The total equipment cost then may range from $9,000 - $50,000. These prices do not include film, processing chemicals, site and storage preparation, and labor. A company or agency may still feel the cost of filming is justifiable when converting a very large paper source archive to microform. Our comparison microform system does include a planetary camera to make it as completely parallel as possible to the optical disk system.

The retrieval of optical disk data is compact, consisting of a disk drive, disk archive, and the same computer and CRT used for input and query. The microform system requires another device--the microfilm reader. This additional piece of equipment is large, containing a viewing screen, auto-scanner and an optional printer. In the top-of-the-line system described earlier, a digitizer and robotic film cartridge loader add to the size and complexity of this bulky apparatus.

3.9 MAINTENANCE AND COST OF OPERATIONS

System maintenance and overall cost of operations for both microform and optical disk storage and retrieval systems is a major consideration. While no absolute quantitative comparisons can be made, it has generally been found, based on discussions with microform system users, that a major criticism of their systems has been related to high cost of system maintenance and operations. System down times have been found to be substantially higher than expected. This is generally due to the mechanical nature of microform systems which results in reduced mean times between failures (MTBF) along with longer repair time.

The electronic nature of optical disk storage systems has been found to be more reliable. Once installed and tested, an optical disk storage and retrieval system requires very little maintenance. MTBF of individual system components generally exceed 30,000 hours. Where failures do occur, simple component or module replacements can quickly be made resulting in minimal system down time. While optical storage systems are still new and therefore do not share the large installed base of microform systems, the electronic components which make up these systems are highly reliable and it can be expected that significantly lower maintenance costs and overall cost of operations can be realized.

SECTION 4. COST

This section compares the latest cost estimates for the optical and microform systems described in section 2.

Costs can vary widely depending on the assumptions made about the configuration and capabilities of the system. A minimum digital system, for example, would consist of a single workstation that would both scan documents and retrieve documents based on index terms and/or abstracts. This would be comparable to the capabilities of a CAR microform system.

A minimum digital system that would also perform full-text search and retrieval would add a recognition processor and full-text software.

Table 3 shows three main examples with two variations of the full-text search example, as follows:

- an image only archive system that does not include full-text search
- an image plus full-text system that includes full-text search and retrieval. Two variations of this are shown, one with index terms and abstracts in addition to the full-text, the other with only minimum manual indexing (title, author, date) relying mainly on full-text.
- A CAR microfilm example assuming outside service bureau film developing, a 16-bit microprocessor control computer, and a reader/printer with auto-scanner.

Table 3 shows nonrecurring costs between the digital systems differing by $20,000. This is the cost of providing full-text search capability--essentially the ability to

Table 3 - Representative Cost Examples

Nonrecurring Cost	Image Only Archive Digital System	Image and Full-text Digital System		Microfilm CAR System
Minimum Workstation	52,000	72,000		58,000
Labor Cost (Recurring) per 1,000 Pages				
Assume:				
pages/document	5	5	5	5
$/hr labor rate	18	18	18	18
pages/hr input	300	300	300	1500
docs/hr indexing	5	5	30	5
docs/hr abstracting	4	5	0	4
pages/hr editing		20	20	
Hours to index	40	40	7	40
Hours to abstract	50	40	0	50
Hours to edit		50	50	
Cost to index	720	720	120	720
Cost to abstract	900	720	0	900
Cost to edit		900	900	
Cost to input docs	60	60	60	12
Total Labor Cost per 1,000 Pages	$1,680	$2,400	$1,080	$1,632

retrieve pages within a document. Labor costs for the examples show the cost for indexing and archiving the digital system in the first column, the cost for indexing, archiving, and editing the full-text conversion in the second column, the cost of minimal indexing and editing the full-text conversion in the third column, and the cost for indexing and archiving the microform system in the fourth column. In the first column it is assumed that more thorough abstracting is done because there will not be any full-text capability to back it up - hence a rate of 4 abstracts per hour rather than 5. In the third column it is assumed that a minimum of indexing including title, author, date, and type of document will be needed in addition to the full-text index. Two minutes per document are allowed for the entry of this basic material.

Table 3 also shows that the labor costs for putting the documents into the database depend more on the level of indexing than on the type of archive media. Columns 1 and 4 show essentially the same labor costs, where the examples for columns 1 and 4 involve thorough manual indexing and abstracting of the documents; column 3 involves only a minimum of separate indices but includes the full-text index. The lower indexing cost and absence of abstracting cost more than offset the effort to verify and correct (edit) text after conversion to machine-readable form. Column 2 shows higher labor cost because that example assumes the documents are indexed and abstracted and that full-text indexing is done as well.

Other examples based on other assumptions are possible using the model shown in table 3. The assumptions can be modified as desired and the new example can be worked using the pattern presented in table 3.

SECTION 5.0 CONCLUSION

This section summarizes the most significant comparisons made in sections 2,3, and 4 of this report.

In reaching a conclusion concerning optical disks versus microform, the following factors should be considered:

- Simplicity and control of archiving process: photo and develop versus scan
- Retrievability of information
- Indexing and retrieval
- Interfacing with digital systems
- Expansion to larger capacities
- Image quality
- Cost

5.1 SIMPLICITY AND CONTROL OF ARCHIVING PROCESS

This consideration tends to favor digital systems. Scanning can be done quickly and conveniently by in-house staff using a clean, safe, dry process. The documents need never leave the premises. Extra document control and tracking procedures are not necessary (as would be needed if documents were sent to a remote location for processing). In the microform process, chemical processes are used to develop the film. Typically, either the developing or the photographing and developing would be sent out to a service center, although with proper procedures both could be done in-house. In the digital system, the entire process is under the control of the data management system. In the microform process, the preparation and storage of the film is separate from the computer assisted retrieval system.

5.2 RETRIEVABILITY OF INFORMATION

This consideration also favors the digital system. The search time to find a desired page, given that it is mounted either on an optical drive or on an auto scanner, is typically under 10 seconds for the optical drive and under 30 seconds for the auto scanner. However, the number of pages accessible without turning a disk or changing film cartridge is 4,000 for a 5-1/4" disk and 5,000 for a 100 foot microfilm cartridge. If a 12" disk were used, 15,000 or more pages could be searched without turning the disk. It is assumed that turning a disk over or changing it completely would be as difficult as removing and inserting a new film cartridge.

5.3 INDEXING AND RETRIEVAL

Index terms and abstracts assist the retrieval of documents. After the documents are located, the pages must be browsed to find the particular items of interest. If the documents are small, the required browsing may not be burdensome. If they are large, full-text search and retrieval may be desired; full-text provides page-level retrieval, which reduces the browsing to pages on which the desired terms are known to appear. Microform systems can only provide document level retrieval. Digital systems can provide page level retrieval.

5.4 INTERFACING WITH DIGITAL SYSTEMS

If the archive system is intended to interface with a digital system for further processing or for transmission to a remote location, the choice is clearly in favor of the optical disk system, where the information is already in digital form.

5.5 EXPANSION TO LARGER CAPACITIES

Both systems can be expanded to larger capacities. The microform system can be configured with robotic retrieval mechanisms that will seek and retrieve film cartridges mounted in special machine-readable racks. Upper limits on the capacity of such machines approach 15,000,000 pages. Prices climb (into 6 figures) steeply as the capacity increases.

Optical disk systems can be expanded by "daisy-chaining" disk drives or by installing a "jukebox". A 50-platter jukebox has a storage capacity of 160 gigabytes (1,500,000 or more pages) at a cost of about $100,000.

5.6 IMAGE QUALITY

Excellent image quality can be achieved in both types of systems. However, special attention must be given to the quality of the microform camera and developer. Many low-end systems produce relatively low quality images because they have not used sufficiently high quality equipment. Also, operator skill is a factor in final film image quality.

Digital scanners operate at 300-400 dots per inch which is in the range of resolution of cameras available for microform systems.

5.7 COST

As shown in table 2, comparable mid-range systems cost about the same. The main consideration is whether full-text will be used, and if so, how much supplementary indexing will be done.

In selecting between optical disk and microform systems, the main decision points are:

- whether full-text search is needed
- whether further processing is intended in any digital system
- whether frequent changing of the film cartridge is a disadvantage during retrieval
- whether the clean, dry, on-site optical capture process is preferred over the microfilm capture process

Other aspects of the system design can be made essentially equivalent.