

2

DTIC  
ELECTE  
NO. 20 1591



# CONSIDERATIONS FOR CONVERSION OF MICROFICHE TO OPTICAL STORAGE

John D. Fowler  
and  
Robert W. Clipper, Jr.

## March 1991

**Thesis Advisor:**

Barry A. Frew

Approved for public release; distribution is unlimited

91-16024



09-16809-2

REPORT DOCUMENTATION PAGE			
1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b RESTRICTIVE MARKINGS	
2a SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b DECLASSIFICATION/DOWNGRADING SCHEDULE			
4 PERFORMING ORGANIZATION REPORT NUMBER(S)		5 MONITORING ORGANIZATION REPORT NUMBER(S)	
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b OFFICE SYMBOL (If applicable) 37	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School	
6c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000		7b ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	
8a NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c ADDRESS (City, State, and ZIP Code)		10 SOURCE OF FUNDING NUMBERS	
		Program Element No	Project No
		Task No	Work Unit Accession Number
11 TITLE (Include Security Classification) CONSIDERATIONS FOR CONVERSION FROM MICROFICHE TO OPTICAL STORAGE			
12 PERSONAL AUTHOR(S) Fowler, John D. and Clipper, Robert W., Jr.			
13a TYPE OF REPORT Master's Thesis	13b TIME COVERED From To	14 DATE OF REPORT (year, month, day) 1991 March	15 PAGE COUNT 156
16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
17 COSATI CODES		18. SUBJECT TERMS (continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUBGROUP	
		Optical Storage, CD-ROM, WORM, Image Scanners, OCR, ICR, Full-text Retrieval, Microfiche, Economic Analysis, Backfile Conversion	
19. ABSTRACT (continue on reverse if necessary and identify by block number)			
<p>Information maintained in a digital format which enables full-text search and retrieval capabilities provides significant advantages over hard-copy or microfiche based information. Recent advances in optical storage technology and full-text retrieval software have made it possible to maintain and access large information bases quickly and inexpensively. Conversion from existing microfiche format to an optical format is possible, although still quite expensive. A case study concerning conversion of microfiche from the Naval Postgraduate School Knox Library Research Reports Division is presented to demonstrate the costs and benefits of having information in a full-text format. All alternatives to full backfile conversion are presented with policy recommendations for organizations considering implementing optical storage systems.</p>			
20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS REPORT <input type="checkbox"/> DTIC USERS		21 ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL Barry A. Frew		22b TELEPHONE (Include Area code) (408) 646-2392	22c OFFICE SYMBOL Code 54Fw

Approved for public release; distribution is unlimited.

Considerations for Conversion of  
Microfiche to Optical Storage

by

John D. Fowler  
Commander, United States Navy  
B.S., Duke University, 1975

and

Robert W. Clipper, Jr.  
Lieutenant, Medical Service Corps, United States Navy  
M.A., University of West Florida, 1983

Submitted in partial fulfillment  
of the requirements for the degree of

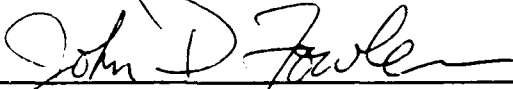
MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

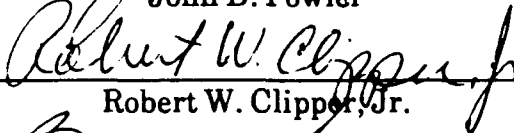
NAVAL POSTGRADUATE SCHOOL

March 1991

Authors:

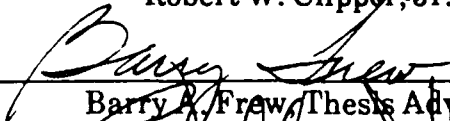


John D. Fowler

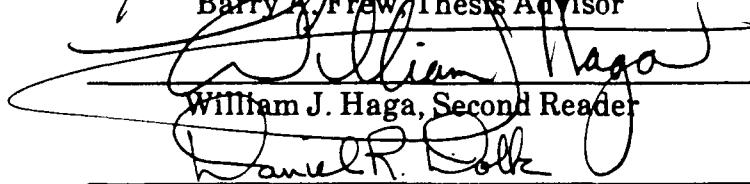


Robert W. Clipper, Jr.

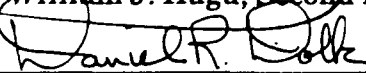
Approved by:



Barry A. Frew, Thesis Advisor



William J. Haga, Second Reader



David R. Whipple, Chairman  
Department of Administrative Sciences

## ABSTRACT

Information maintained in a digital format which enables full-text search and retrieval capabilities provides significant advantages over hard-copy or microfiche based information. Recent advances in optical storage technology and full-text retrieval software have made it possible to maintain and access large information bases quickly and inexpensively. Conversion from existing microfiche format to an optical format is possible, but still quite expensive. A case study concerning conversion of microfiche from the Naval Postgraduate School Knox Library Research Reports Division is presented to demonstrate the costs and benefits of having information in a full-text format. Alternatives to full backfile conversion are presented with policy recommendations for organizations considering implementing optical storage systems.

Accession For	
NTIS ORAL	<input checked="" type="checkbox"/>
NTIS TAB	<input type="checkbox"/>
Accession Used	<input type="checkbox"/>
Justification	
By	
Distribution	
Availability Codes	
Avail and/or	
Dist	Special
A-1	

## TABLE OF CONTENTS

I.	INTRODUCTION . . . . .	1
	A. INFORMATION OVERLOAD . . . . .	1
	B. METHODOLOGY . . . . .	3
	C. CATEGORIES OF INFORMATION . . . . .	3
	D. ORIGINAL PAPER SOURCE DOCUMENTS . . . . .	6
	E. MICROFORM . . . . .	7
	F. OPTICAL STORAGE SYSTEMS . . . . .	12
	G. ADVANTAGES OF OPTICAL MEDIA . . . . .	15
	H. DISADVANTAGES OF OPTICAL MEDIA . . . . .	18
	I. MAJOR ISSUES IN CONVERSION TO OPTICAL STORAGE . . . . .	19
II.	BACKGROUND/HISTORY . . . . .	21
	A. INFORMATION STORAGE AND RETRIEVAL SYSTEMS . . .	21
	B. THE MEMEX . . . . .	25
	C. ON-LINE INFORMATION SERVICES . . . . .	27
	D. ON-LINE, INTERACTIVE, DIGITAL INFORMATION SYSTEMS . . . . .	27
	E. THE MEMEX TODAY . . . . .	30
III.	OPTICAL MEMORY SYSTEMS . . . . .	32
	A. OPTICAL DISC STANDARDS UPDATE . . . . .	32
	B. RECORDING AND READING TECHNIQUES FOR OPTICAL MEDIA . . . . .	35
	C. COMPACT DISC READ ONLY MEMORY (CD-ROM) . . . .	36
	D. WRITE ONCE READ MANY (WORM) . . . . .	41
	E. OTHER TYPES OF OPTICAL DISC . . . . .	42

IV.	OPTICAL DATA-ACQUISITION SYSTEMS . . . . .	44
A.	OPTICAL SCANNING HARDWARE . . . . .	44
B.	IMAGE SCANNING . . . . .	47
C.	OPTICAL CHARACTER RECOGNITION (OCR) . . . . .	50
D.	PERFORMANCE CHARACTERISTICS . . . . .	53
V.	DOCUMENT RETRIEVAL SYSTEMS . . . . .	55
A.	DOCUMENT RETRIEVAL DEFINED . . . . .	55
B.	IMAGE-BASED VERSUS TEXT-BASED STORAGE . . . . .	58
C.	IMAGE-BASED SYSTEMS . . . . .	58
D.	TEXT-BASED SYSTEMS . . . . .	59
E.	THREE TYPES OF ELECTRONIC DOCUMENT RETRIEVAL SYSTEMS . . . . .	61
F.	RETRIEVAL SOFTWARE FEATURES . . . . .	63
G.	SELECTION CRITERIA . . . . .	67
H.	IMPORTANCE OF RETRIEVAL SYSTEMS . . . . .	67
VI.	TECHNOLOGY FOR MIGRATING IMAGES TO OPTICAL DISC BASED SYSTEMS . . . . .	69
A.	THE NEED FOR IMAGE MIGRATION . . . . .	69
B.	EARLY RESEARCH INTO MICROFORM SCANNING . . . . .	70
C.	MICROGRAPHICS TO OPTICAL CONVERSIONS IN PROGRESS . . . . .	72
D.	READER-PRINTERS AND READER-SCANNERS . . . . .	75
E.	MICROFORM IMAGE SCANNERS . . . . .	76
F.	THE FEASIBILITY OF MICROFORM DIGITIZATION . . . . .	79
VII.	ANALYSIS OF THE REQUIREMENTS FOR MIGRATING UNCLASSIFIED TECHNICAL REPORTS FROM MICROFICHE TO OPTICAL DISC IN THE KNOX LIBRARY, RESEARCH REPORTS DIVISION . . . . .	81
A.	METHODOLOGY . . . . .	81

B.	REQUIREMENTS ANALYSIS . . . . .	81
C.	DESCRIPTION OF REQUIREMENTS THAT ARE NOT BEING MET BY THE CURRENT SYSTEM . . . . .	87
D.	COMPATIBILITY LIMITED REQUIREMENTS . . . . .	90
E.	RECORDS MANAGEMENT REGULATIONS . . . . .	94
F.	TRAINING REQUIREMENTS . . . . .	95
G.	SPACE AND ENVIRONMENT . . . . .	96
H.	CAPABILITY AND PERFORMANCE VALIDATION . . . . .	97
I.	SUMMARY OF REQUIREMENTS . . . . .	99
VIII.	ANALYSIS OF ALTERNATIVES FOR MIGRATING UNCLASSIFIED TECHNICAL REPORTS FROM MICROFICHE TO OPTICAL DISC IN THE KNOX LIBRARY'S RESEARCH REPORTS DIVISION . . . . .	101
A.	THE NEED FOR AN ANALYSIS OF ALTERNATIVES . . . . .	101
B.	THE SIZE AND SCOPE OF THE ANALYSIS . . . . .	102
C.	INFORMATION OBTAINED CONCERNING THE MARKETPLACE . . . . .	103
D.	IDENTIFICATION OF THE ALTERNATIVES . . . . .	106
E.	DETERMINING THE MOST ADVANTAGEOUS ALTERNATIVE . . . . .	117
F.	THE DECISION PROCESS: SELECTING AND REPORTING THE MOST BENEFICIAL ALTERNATIVE . . . . .	121
G.	A DISCUSSION OF THE RESULTS OF THE ECONOMIC ANALYSIS . . . . .	126
IX.	CONCLUSION AND RECOMMENDATIONS . . . . .	132
A.	CONVERSION TO FULL-TEXT FORMAT IS POSSIBLE . . . . .	132
B.	THE DISCIPLINE OF ECONOMIC ANALYSIS SHOULD BE USED . . . . .	132
C.	KNOX LIBRARY RESEARCH REPORTS DIVISION RECOMMENDATIONS . . . . .	133
D.	CONCLUSION . . . . .	135

APPENDIX A	CHECKLIST FOR ASSESSING SOFTWARE RETRIEVAL CAPABILITIES . . . . .	137
APPENDIX B	REGULATIONS FOR INFORMATION RESOURCE MANAGEMENT . . . . .	139
APPENDIX C	VENDORS REPLYING TO THE COMMERCE BUSINESS DAILY ADVERTISEMENT . . . . .	141
APPENDIX D	LIBRARY SYSTEMS EXPERT OPINION SURVEY . . . . .	142
	LIST OF REFERENCES . . . . .	143
	INITIAL DISTRIBUTION LIST . . . . .	147



## I. INTRODUCTION

### A. INFORMATION OVERLOAD

Information is accumulating around us at an ever increasing rate (Naisbitt, 1982). The three associated problems of storing, cataloging, and retrieving vast amounts of information present a formidable challenge to records managers, librarians, researchers and any one who must handle the deluge of information available today. These three problems are inherently intertwined; the medium used for storage influences the means of cataloging, and this in turn influences the method of information retrieval. The growing masses of information we encounter in our daily lives has caused us to be unable to effectively deal with the overload. This overload is a by-product of a shift from an industrial-based economy to one that is information-based.

This "information revolution...is momentarily stalled for want of easy, intelligent access to the masses of data we are accumulating" (Toffler, 1981). The problem is no longer a lack of information, rather it is an inability to deal with the "glut of unrefined, undigested information flowing in from every medium around us" (Toffler, 1981).

This is not a new problem; it was recognized in July of 1945 by the Director of the Office of Scientific Research and Development, Vannevar Bush. In an article in Atlantic Monthly, July 1945, he described the problem as a growing mountain of data that is expanding beyond man's capability to handle effectively. Specialization, Bush notes, has caused an increasing proliferation of information. While our ability to publish this information has kept pace with the trend, our ability to navigate through such vast quantities of information has lagged far behind. Bush describes the plight of the researcher as follows:

The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships. (Bush, 1945)

In spite of the progress made since 1945, advances in technology only serve to hold our position steady in relation to the accelerating growth of information.

Advances in microform technology have enabled us to increase the compression factor from 20 to 1 in 1945 to accepted standards of 24, 42 or 48 to 1, with 96 to 1 factors available in experimental applications (Saffaday, 1978). This limited progress has enabled us to deal with the issue of storing vast quantities of information more compactly, but does nothing for the associated issues of cataloging and retrieval. Vannevar Bush predicted these advances in microform technology while acknowledging that

they did not address the more important issue of effectively distilling the information. He realized that the ability simply to retrieve information was not enough; one needed the ability to selectively filter the information.

The problem of filtering, and related issues concerning cataloging and retrieving are even more important today. Given the growth of recorded information, we need an effective means of selectively accessing the required information. A computer-based information system is essential to automate the access. However, this technique by no means answers all of the concerns inherent in the problem.

## **B. METHODOLOGY**

In preparation for a discussion of the advantages and disadvantages of automating an information system, we will present the essential issues. Background and technology associated with storing, cataloging, and retrieving information will be presented first, followed by a case study which will apply these technologies to the Research Reports Division (RRD) of the Naval Postgraduate School Knox Library.

## **C. CATEGORIES OF INFORMATION**

A distinction must be made between three categories of information encountered in information systems. The type of

information and its primary use will determine the appropriate storage medium to be used. The three types are as follows:

- computer-based information,
- draft information, and
- document-based information.

### **1. Computer-based Information**

Computer-based information is valued for its timeliness and accuracy. It consists of temporary, or working information that is designed to be changed regularly. Two examples are databases of employee phone numbers and working spreadsheets of quarterly income and expenses.

Rewritable media provide an easy modification capability by overwriting the existing data. Therefore, magnetic media, such as Winchester disks, are most appropriate for computer-based information. Because computer-based information is not intended to be stored for long periods of time, it will be excluded from our study.

### **2. Draft-based Information**

Draft-based information is information created on word-processors or similar software that is not yet in final form. This information derives its greatest value from being modifiable as it is intended to be used again for

changes and additions. Early iterations of memoranda, letters, instructions, and notices are good examples of draft-based information.

Rewritable media is also appropriate for draft-based information because it is easily modifiable. Accordingly, draft-based information will also be excluded from our study.

### **3. Document-based Information**

Document-based information comprises the third category of information and accounts for more than 90 percent of all information in today's offices. (Toffler, 1981) This type of information provides a formal, unmodifiable record for reference, transaction, and evidentiary purposes and will be the focus of our study. The six features of a formal document are listed below.

- The originator must be clearly identified.
- The recipient must be identified.
- It must be dated or dated and timed.
- It must show the approving signature or initials.
- It must be a complete and final entity.
- It must be sealed after approval. Changes can only be made with the originator's approval. (Waegemann, 1989)

Rewritable media are decidedly not appropriate for document-based information. The issue of identifying the originator and verifying his signature can be handled today via

biometrics such as a retinal scanner or a thumb scanner attached to the user's computer. These scanners require positive identification prior to storing a document, but sealing of a magnetic-media document after signature is impossible to implement.

Easy modification via overwrite inherent to magnetic media is an impediment to its use as an archival medium for document-based information. Therefore, some other medium must be chosen for the storage of document-based information.

Traditional alternatives of original paper source documents and microform have only recently (since 1985) been joined with computer-based optical storage systems. In the following section, each alternative and its accompanying advantages and disadvantages will be examined.

#### **D. ORIGINAL PAPER SOURCE DOCUMENTS**

##### **1. Paper Advantages**

Advantages of paper storage are readily apparent, but often taken for granted. Three advantages of original source documents are listed below:

- non-modifiable, (any attempt to alter the original will be apparent);
- available, (no conversion costs are required); and
- traditionally accepted as evidence, (no legal challenges are to be expected).

## **2. Paper Disadvantages**

The disadvantages of paper are less apparent and are often overlooked. The disadvantages of original source documents are described below.

- The cost of accessing a document, (including the cost of: accessing the equipment - file cabinet, accessing the container - file folder, referencing and inserting the document, restoring the container, restoring the equipment, and returning to the work place. (Waegemann, 1989))
- The cost of the storage space for the documents (2000 pages occupy about one linear filing foot of space).
- The non-availability cost of the paper document. (This is the cost attributable to not having a document available when needed.)

For relatively small document storage systems, where risk exposure to non-availability of documents is low, a paper-based filing system may be the most economical.

## **E. MICROFORM**

As the volume of paper-based information increases past an organization's ability to manage it effectively, other solutions must be sought. A traditional answer to the problem of how to store this accumulating record of information has been to put it on microform. Microform is the generic term which includes microfilm (reel or cassette), microfiche (4x6 inch sheets), and aperture cards (computer punch cards with a small section of microfilm inserted in a cutout in each card).

## **1. Microform Advantages**

The primary advantages of microform over the original source documents are listed below.

- Microform requires (much) less space. A standard 4 by 6 inch microfiche can contain 98 images at a compression ratio of 24 to 1.
- Microform is far lighter and therefore cheaper to mail.
- Microform provides unitization - it groups records together in a fixed sequence so individual records won't be misplaced.
- Microform documents are more durable and require less careful handling than originals.

Listed below are advantages that microform has in common with original source documents.

- Microform images are unalterable (any tampering with the images would be detected).
- Individual microform images cannot be deleted (short of destroying an entire sheet).

For the reasons above, microform is well suited to its traditional role as the archival medium of choice for records managers.

## **2. Microform Disadvantages**

The primary disadvantages of microform over original paper source documents are noted below.

- Microform storage incurs conversion costs to photograph the images of the original documents.



- Microform storage requires the use of microfiche or microfilm readers to view the stored documents.
- Microform storage requires the use of microform reader/printers to convert the document back to hard copy.
- Microform storage is inconvenient and awkward for users to access.

For the primary advantage of obtaining more compact (smaller and lighter) storage, microform incurs additional costs in terms of hardware and retrieval time. Since the hardware costs are modest and can be amortized over many retrieval operations, these costs do not create a significant barrier to the use of microform. However, the issue of retrieval is significant and it has been addressed through automating the microform retrieval process.

### **3. Computer Assisted Retrieval (CAR)**

#### **a. General**

Computer assisted retrieval systems involve manually indexing microform documents, maintaining an automated index, and using a computer-based automatic retrieval system to locate a particular microform image.

#### **b. Microfilm Retrieval Systems**

Microfilm retrieval systems require frame locating "blips" containing an index number to be inserted with each frame as it is photographed, or an optical frame counting device attached to the microfilm reader. In either case, an index which matches key document identifiers with

reel and frame location numbers is built and maintained. To retrieve a document, the user issues a query for the document title, whereupon the system index responds with a cassette number. The user is prompted to install the appropriate cassette and the cassette is driven to the appropriate frame number. The user can then view or print the desired document on the associated microfilm reader-printer.

#### **c. Microfiche Retrieval System**

A microfiche computer assisted retrieval system operates on the same principles as a microfilm retrieval system except that in place of a motor driven microfilm reader-printer, there is a motor-driven microfiche cartridge reader-printer that holds a group of microfiche. When the user queries the index for a document title, the system responds with a cartridge number. The user is then prompted to install the appropriate cartridge whereupon the cartridge selects and positions the desired image on the microfiche reader-printer. The user can then view or print the desired document.

#### **d. Aperture Cards**

Aperture cards are also a form of computer assisted retrieval. The microform images contained in the punch card cutouts are indexed by a keypunch operator. The space for indexing on an aperture card is limited because

only 59 of the 80 columns are available for encoding after the microform image has been inserted. When an image is requested, the system locates the desired aperture card and loads it into the microfilm reader printer for display or printing.

#### **e. Summary**

Computer assisted retrieval offers the user the option of trading increased CAR hardware and software cost for the increase in accessibility achieved through reducing retrieval time. It applies the advantages of computerized indexing, search, and retrieval to the established microform technology.

#### **4. Microform/Paper Similarities**

Microform and paper both treat the document as the smallest retrievable unit in the system. In order to obtain information from within a document, the user must retrieve and read the document. Additionally, in order to access the document, the user must know the key terms used to index the document (i.e., the name of the file folder). The user can only access documents via those keys that are "known" to the index. If he attempts to search on a key that has not been indexed, his search will be unsuccessful. For example, unless a document pertaining to CD-ROM is indexed under "optical storage", it will be invisible to a user who

consults an index for all documents on optical storage. This characteristic presents a significant limitation.

Paper-based storage, microform-based storage, and optical storage form a continuum progressing from the least to the most automated information storage systems. For this reason it would be of little use to compare optical storage with paper. We will, however, compare optical storage systems with microform storage systems as we investigate the feasibility of converting from microform to optical disk.

#### **F. OPTICAL STORAGE SYSTEMS**

We will discuss three major functional divisions in optical storage and their strengths and weaknesses with respect to document-based information storage (archives). The three functional optical storage categories discussed are as follows:

- Compact Disc - Read Only Memory (CD-ROM),
- Write Once Read Many (WORM), and
- erasable optical media.

##### **1. Compact Disc - Read Only Memory (CD-ROM)**

CD-ROM is an optical storage medium which is derived directly from the technology of Compact Disc - Audio. The most significant feature of CD-ROM is its ability to store over 540 megabytes (MB) of data on a single 4.72 inch diameter disc (Lambert and Ropiequet, 1986). This is the

equivalent of over 1250 low density floppy disks or 450 high density, 1.2 megabyte disks. This ability to store extremely large quantities of data has made CD-ROM an excellent choice for archiving information under certain circumstances. Because of the high fixed costs associated with "pressing" a CD-ROM disk it is primarily a distribution or publishing medium. However, if there is a requirement for multiple copies of a large body of data, economies of scale quickly come into play and make CD-ROM competitive with other forms of mass storage. CD-ROM's major disadvantage is a product of its CD-Audio heritage.

The same characteristics that enable the dense packing of information on a disk hinder the quick retrieval of that information. Information retrieval times of CD-ROM are considerably greater than those of magnetic media, but for a well-designed application it is still less than a second. As its name implies, CD-ROM is a read-only medium. This means that there is no overwrite capability. While this may initially appear to be a disadvantage to the computer user who is familiar with magnetic media, it definitely is not a disadvantage for certain types of information.

It is essential that archival, catalogue, and regulatory information not be altered and therefore the absence of an overwrite capability in CD-ROM gives the

information assured permanence and this optical storage medium a decided advantage for these applications.

Another distinct advantage of CD-ROM is the existence of standards. International Standards Organization standard 9660 sets specifications for the physical and logical requirements for information on a CD-ROM. The availability of standards ensures that a CD-ROM manufactured by one company is readable on any other ISO-9660 compatible CD-ROM drive. This portability of data is a great advantage especially in an open systems environment that is likely to exist in the future.

## **2. Write Once Read Many (WORM)**

WORM discs have many of the same advantages as CD-ROM discs which are; a very dense storage capability (up to 600 megabytes on a 5.25 inch disc) and the absence of an overwrite capability (Waegemann, 1989). The WORM disc therefore qualifies as an appropriate archival medium. Another advantage of WORM is the ability to write directly to disc without having to send information to an outside source for disc production.

The major disadvantage of a WORM disc when compared with CD-ROM disc is the higher unit cost. A formatted WORM disc can cost from \$100 to \$200 each, whereas a CD-ROM disc can be as inexpensive as \$2 when produced in volume. WORM discs now have a standard (ISO 9771) which means portability

of WORM discs among WORM drives. For a single-site information management system, a WORM drive option may be the most economical optical storage system.

### **3. Erasable Optical Media**

Erasable optical technology has many of the best characteristics of optical and magnetic media. It provides a high density, high capacity storage medium with the ability to overwrite information no longer current or desired. When improvements in the speed of access time and establishment of industry standards are developed, erasable optical media will be in competition with current magnetic media. However, the existence of an overwrite capability renders it inappropriate as an archival medium and therefore it will not be addressed in depth in this study.

## **G. ADVANTAGES OF OPTICAL MEDIA**

Optical media suitable for archiving document-based data include CD-ROM and WORM. These media have three very significant advantages over microform: compactness, unitization, and an on-line, digital format.

### **1. Compactness**

CD-ROM surpasses the compression factor of standard microform by a factor of 40 in terms of weight. (Lind, 1987) This becomes particularly important if distribution of data is a consideration. The advantages of being able to put so much information onto such a small disc are significant, but

not sufficient to justify conversion to optical media. A reduction in media access time will yield reductions in cost but probably not be sufficient to offset increased system acquisition costs.

## **2. Unitization**

Optical media goes well beyond the unitization capability of microform by permitting 540 MB of information on one disc. This feature reduces, if not eliminates, the problem of misfiling or losing documents (Lambert and Ropiequet, 1986). Unitization, putting an entire information base on one disc, has advantages beyond the obvious one of being unable to lose or misfile a record. The fact that all information resides permanently in its own location on the disc means that no refiling costs are ever incurred. Only a copy of the information is actually provided to the user so it need not be replaced. The biggest advantage of unitization is that it guarantees 100 per cent record availability.

## **3. On-line, Digital Format**

Optical media store information in an on-line, digital format. This has several significant implications for storage systems that it can support. On-line media can support character-based as well as image-based systems, direct manipulation of text, graphics output, and full-text retrieval of information. The implications of this on-line



capability give optical media a clear advantage over microform.

**a. Image-based versus Text-based Systems**

One advantage that accrues to text-based systems is that of information density. When documents are stored as images in digital form, even after data compression, they occupy considerably more space than the same documents stored in an ASCII coded format. For example, a document takes approximately 25 times more space when stored as a 300 dot per inch raster scanned image than when stored as text (Navy Publications and Printing Service, 1990). To further illustrate the storage savings of text-based systems consider that a typical CD-ROM disc can hold about 270,000 documents in text form compared to 10,800 in image form.

**b. Direct Manipulation of Text**

Having documents stored in a text-based format makes it possible to copy the text into other documents for word-processing purposes. For applications where the information contained in the documents is to be merged or combined with other text, this is a very significant advantage. This capability is not available in an image-based system.

**c. Graphics Presentation**

Data extracted from documents can be presented in graphics format if desired, provided the system is text-

based. For example, if a document contained data on net sales versus advertising expenses, this information could be extracted and entered into a graphics program that could provide a visual display of the relationship between the two, rather than simply present the raw data. This has important implications for reducing the quantity of data that must be analyzed by a decision maker or researcher and enhances the usefulness of the data.

#### **d. Full-text Retrieval**

One of the major advantages of on-line digital systems is the ability to store text-based information rather than only image-based information. The distinction is one primarily of granularity; of the size of the smallest addressable unit of the information base. In a text-based system each word in the system is addressable while in an image-based system, the smallest addressable unit is the document. A text-based system has intelligent documents which can be queried for content. An image-based system on the other hand has non-intelligent documents which permit no such queries based on their content. The ability to search a document for words or combinations of words is known as "full-text retrieval" and is a very powerful advantage.

#### **H. DISADVANTAGES OF OPTICAL MEDIA**

The primary disadvantage of optical media lies in the conversion costs for existing systems. The improvements in

optical scanning and intelligent character recognition have made conversion possible, however it is expensive. While scanning and character recognition are automated processes, they still require human intervention to perform quality assurance and problem resolution. Converting to image-based optical systems where automatic indexing is not possible includes a cost for manually entering key index fields. This can be a substantial cost. For converting to character-based systems, automatic indexing software exists and can reduce some of the human effort required.

#### **I. MAJOR ISSUES IN CONVERSION TO OPTICAL STORAGE**

The three major issues to be resolved when converting microform to optical storage systems are as follows:

- acquisition systems, (conversion of the information from microform to digital)
- storage systems, (determination of storage media) and,
- retrieval systems, (cataloging or indexing, and retrieval of the information once on the optical medium).

Each of these issues will be addressed in detail in subsequent chapters.

##### **1. Summary**

We have introduced the concept of document-based information and its archival nature and have demonstrated the need for a permanent, non-alterable medium for this type

of storage. We have thus ruled out magnetic media as well as erasable optical media and are left with five possibilities for archival storage. The possibilities and their accompanying traits are as follows:

- Paper-based original source documents which are expensive due to space maintenance, and non-availability costs.
- Microform without Computer Assisted Retrieval which is not feasible for large systems because of long retrieval times.
- Microform with Computer Assisted Retrieval which is feasible but expensive. In addition, it is a lagging technology which only postpones the conversion decision.
- Conversion to CD-ROM, having a high initial unit cost, can be reduced significantly as economies of scale are encountered through replication and distribution of multiple copies of the database.
- Conversion to WORM which has a somewhat lower initial unit cost than CD-ROM and is economical for single site applications.

The first three alternatives all have significant shortcomings that render them less than optimal for future information storage and therefore emphasis will be placed on CD-ROM and WORM systems. After discussing acquisition systems, optical storage systems, and retrieval systems in detail, we will examine the technology required for conversion to optical storage and present an analysis of alternatives for the specific case of the Knox Library RRD microfiche collection.

## **II. BACKGROUND/HISTORY**

### **A. INFORMATION STORAGE AND RETRIEVAL ISSUES**

#### **1. Image-based Versus Text-based Systems**

Information system managers must be able to deal with issues of acquisition, storage, and retrieval of information. The intended application of the information influences the best retrieval method, which in turn influences the format in which the information should be stored. The format influences the information acquisition or conversion strategy. The two format options in computer-based systems are image-based storage and text-based storage. The advantages and disadvantages of each format option are discussed in the sections below.

##### **a. Image-based Storage**

If the intended application of an information system is for legal record keeping, or archiving, the integrity of source documents can best be maintained when stored as fully reproducible images. Unfortunately the cost of storing documents in image form is significant. For example, a single CD-ROM disc can hold only 10,800 document images compressed to 50 kilobytes each but when the information is stored as ASCII-coded text, it can hold 270,000 pages of 2000 character documents. An advantage of image-based storage is that the technology required to

convert a paper or microform document to an image is far less complicated than that required to convert that image into text. While there is not yet a legal precedent establishing the admissability of computer based images as evidence in court, a ruling is expected. Establishment of such a precedent would aid the overall acceptance of image-based optical storage systems.

**b. Text-based Storage**

If the intended application requires a search for, and extraction of, information from within documents then a text-based system will be more useful. As shown above, the capacity of a text-based system is far greater than a comparable image-based system, and a text-based system provides increased functionality by permitting a full-text search capability. However, these added capabilities come at a price. If the documents must be converted from a microform or paper-based system, an Optical Character Recognition (OCR) or an Intelligent Character Recognition (ICR) process will be required to convert a scanned image into the corresponding ASCII-coded text. This type of system may prove to be quite labor intensive, and therefore expensive, especially in the area of quality assurance.

## **2. Retrieval Mechanisms**

### **a. Image-based Retrieval**

When documents are stored as images, there is no access to information stored within a document, therefore an index must be built to identify each image by one or more key words. This type of storage and retrieval is suited to applications where strict archival procedures must be maintained or where all retrieval is document- or transaction-oriented as opposed to information-oriented. For example, where all retrievals from storage were made by name or invoice number, an image-based storage system would be useful.

### **b. Text-based Retrieval**

Access to the information content of a document provides significantly improved retrieval capabilities to the information system. The ability to retrieve all documents that contain the words "CD-ROM" and "retrieval" demonstrates an increased functionality over an image-based retrieval system. This ability to "look within" documents in an information system is particularly useful in research environments where the researcher seeks to increase his knowledge of a given subject. A text-based retrieval system lets him search beyond the limits that might be established by an indexer and permits him to interact with the information contained within each document. Most text-based

retrieval systems offer the same key index features that are possible in an image-based system.

### **3. Acquisition Processes**

Just as the intended application for an information base determines the storage format, the format determines the degree of complexity of the acquisition process. As publishing has become more computerized, advances have been made in automating the acquisition of information in computer-usable forms. Since most publishing is done electronically today, it is possible to obtain the text of a document already in electronic form. If, however, conversion is required from existing microform or paper documents, the use of scanners to digitize the information is necessary for either image-based or text-based systems. A text-based system must take the additional step of converting the digitized image into text. This step requires the additional technologies of optical character recognition or intelligent character recognition. These technologies will be discussed in chapter four. Because the intended application of an information system determines many of its characteristics, we will investigate the requirements for each system and the technologies to support them.



## **B. THE MEMEX**

The idea of developing a system to allow the acquisition, storage, and retrieval of large information bases is not a new one. Vannevar Bush, as noted in the beginning of this paper, not only recognized the problem of information overload, he envisioned a solution to the problem. Except for his use of analog rather than digital information storage techniques, he quite accurately described what we have come to know as the personal computer. His vision is all the more remarkable in view of the fact that the stored program digital computer would not be invented until 1947.

Bush envisioned a device to extend man's ability to deal with the information overload he faced. He called it the Memex. His Memex included a keyboard, a slanting translucent screen, and a section for storage of information. The primary feature of this device was the ability of the user to consult his books, notes, and communications which had been stored in the Memex on microform, with "exceeding speed and flexibility". A Memex owner, in Bush's vision, would be able to buy microform documents that could be read into the Memex. He would be able to retrieve those documents by using the keyboard provided. This description could fit a computerized aperture card system or a CAR microform system, since either allows for automated retrieval of microform images.

However, Bush had revolutionary ideas about how one should be able to manipulate the stored text. In addition to standard indexing, he made the leap to associative indexing. His system demanded that access be provided to the contents of documents - to the ideas they contained.

...associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. The process of tying two items together is the important thing. (Bush, 1945)

Bush has clearly described what Ted Nelson later named HyperText in the 1960s. The ability to follow a train of thought, forward or backward through a body of information is central to such a system. Clearly if we are to realize a capability of "associative indexing" we must be able to address a unit of information smaller than the document. We must be able to focus our attention on a given paragraph or word within a document.

This fine degree of granularity can only be obtained in a system which stores documents in a text-based format. Microform or image-based systems do not provide the ability to see below the document level which is essential to HyperText or "associative indexing". Failure to develop this capability will prevent us from realizing the full value of stored information.

Bush's vision was well ahead of his time. The technology was not yet developed that would enable the

creation of such a machine. It would take the development and commercialization of both the digital computer and inexpensive on-line storage media to make the Memex a reality.

#### **C. ON-LINE INFORMATION SERVICES**

The management of large, dial-up computer databases only became feasible when the combined cost of storing large quantities of information combined with the telecommunications costs became inexpensive enough to make the databases profitable. The great expense of maintaining large dial-up, on-line databases required that the fixed costs of maintaining and operating a mainframe computer and large on-line storage facilities be distributed over a large user base. The existence of large, multi-user information systems led to advances in the acquisition of documents in digital form as well as in retrieval mechanisms. These advances provided the underlying infrastructure that made optical storage feasible. It was not until the 1980s that the emerging optical disc technology made inexpensive, on-line storage of large amounts of information available to anyone with a personal computer.

#### **D. ON-LINE, INTERACTIVE, DIGITAL INFORMATION SYSTEMS**

Optical information systems differ from paper or microform based systems in compactness, degree of unitization, and in the degree of computer control. The

primary advantages of optical systems over microform systems derive from the fact that optical systems can be on-line, interactive, and digital.

### **1. On-line Information Systems**

An on-line system is one that is under the control of a computer; once initiated by a user, it does not require human intervention for its operation. (Sanders, 1983)

Examples of on-line storage are magnetic hard disk drives, reels of magnetic tape installed on tape drives, and CD-ROM discs in CD-ROM drives or multiple disc autochangers or "jukeboxes". Even a reel of microfilm mounted on a computer assisted retrieval (CAR) system could be considered on-line storage. On-line storage allows quick automatic access to information. A CD-ROM system can access a one-page document from a group of 270,000 on a single disc in an average of .5 seconds (Lambert and Ropiequet, 1986).

The limits of an on-line system are encountered when human intervention is required to gain access to data. For example a reel of magnetic tape in the computer center's library, a CD-ROM disc not installed in a drive, and a roll of microfilm not installed on a CAR system are examples of off-line storage.

### **2. Interactive Information Systems**

An interactive system is one in which the user carries on a dialogue with the computer. This is in contrast to a batch system in which the user tells the

computer what series of functions to perform and then waits for the batch process to be executed in order to receive the output. The difference is one of responsiveness. In an interactive system the response time of the system is critical. In accessing information from a database, a user is concerned with a quick, accurate response to his query. Once he has received the response, if he is operating in an interactive mode, he can improve upon the query and move iteratively toward his goal.

### **3. Digital Information Systems**

The digital computer has become so pervasive in our lives that we take the digital aspect of it for granted. We expect any computer-based information system to be able to search its database for words matching a given criteria or to be able to find any combination of words that exists within a document. These functions can only be performed on databases that are stored in digital format that permits string searches of the stored digital codes. This capability distinguishes optical from microform based systems. Because microform based systems are analog in nature, there is no ability to manipulate the text of the images. In any application where it is desirable to work with the text of documents, optical systems have the advantage of being able to store the information digitally in a text format. This makes the text of each document available to the researcher. Optical systems can also store

images in raster form however this is only a minor improvement over the original microform based storage. Images stored on optical media in raster format occupy far more space than those stored as text and do not permit manipulation of the content of the text.

#### **E. THE MEMEX TODAY**

Optical storage combined with digital technology has now extended the on-line, interactive, digital storage available in a personal computer environment to the point where the Memex is quite feasible. Bush envisioned his user to be able to insert up to 5000 pages of text a day into his Memex with no overload problem. If each page contained 2000 characters of text, then 10 megabytes of storage capacity would be needed daily. CD-ROM provides 540 megabytes of storage per disc and WORM provides 600 megabytes per 5.25 inch disc (Wagemann, 1989). Optical storage combined with the ever-increasing power of the micro-computer has made the Memex technologically, operationally, and economically feasible.

It is often said that new technologies are often solutions in search of problems. That is, the technology has been developed, but not a methodology to employ it. In the case of optical storage allied with the micro-computer processing power, we now have the ability to provide rapid access to vast amounts of information that Vannevar Bush could only imagine. Victor Hugo stated that no one can

resist an invasion by an idea whose time has come. Optical storage is just such an idea. Coupled with the developing scanning and recognition techniques now available, and the information retrieval capabilities derived from on-line information services, we have a viable methodology for transferring information from microform storage to optical storage. It will be the medium of the future, and depending on the application, it may be the medium for today.

Advances in the technologies of acquisition, storage, and retrieval of information have progressed to the state where the methodology for transferring information bases to optical storage is viable. Bush's Memex is within our grasp. The following sections will examine the advances in the three areas of information management that have made this possible.

### III. OPTICAL MEMORY SYSTEMS

#### A. OPTICAL DISC STANDARDS UPDATE

The development of standards in emerging technologies may cause a company to lose its original investment if a competing standard is adopted. An example of this was the beta video recording technique.

In the field of optical disc, only Compact Disc-Read Only Memory (CD-ROM) has an established standard that is widely accepted. This standard is composed of a set of specifications defined in the International Standards Organization (ISO) 9660.

The CD-ROM standard is the result of cooperation between the CD-ROM industry leaders including: Apple Computer Company, Digital Equipment Corporation, Hewlett-Packard, Philips, and Sony. The leaders met in 1987 at Lake Tahoe, California to develop CD-ROM standards and are now popularly known as the "High Sierra Group". Their resulting industry cooperative effort is credited with the booming expansion of the CD-ROM market. The basic idea is that if a CD-ROM disc drive meets the ISO 9660 standard then it should be able to use any disc conforming to the standard.

Outside the domain of CD-ROM, the standards issue is yet to be resolved. However, a new set of standards has recently been adopted for 130mm Write Once Read Many (WORM)



drives. These standards are defined in ISO 9171. Many other standards are pending, Table 1 lists those available at this time.

Standards are very desirable from the end-users' perspective. They provide portability of applications and increase the size of potential markets, thereby reducing the costs of new technology. Historically, standards have been difficult to achieve, due to competition among manufacturers.

When manufacturers do achieve establishment of a standard, an economic effect on the market results. Standards increase the supply in the market, increased supply drives the price down, and reduced costs increase the demand until the market is saturated or reaches equilibrium.

CD-Audio is a good example of a standardization success in the marketplace. Sony and Philips corporations agreed on a standard, and were able to increase the supply in the market and reduce the price. Today CD-Audio players and discs are readily available at reasonable prices. CD-ROM, based on the CD-Audio standard may soon be a household word as its momentum in the marketplace increases.

Standards describe the physical and logical format of a disc. For example, CD-ROM discs are addressed by minute, second, and sector. By standardizing on this addressing

**TABLE 1. OPTICAL DISC STANDARDS**

	Millimeters	Inches	Standard	Comments
WRITE ONCE READ MANY (WORM)	356	14	ISO CD-10885	Committee Draft
	300	12	NS	No standard exist
	130	5.25	ISO 9171-1 (1990) ISO 9171-2 (1990)	Part 1: Unrecorded optical disc cartridge Part 2: Recording format
REWRITABLE OPTICAL	130	5.25	DIS 10089	Optical disc cartridges for information inter- change
CD-ROM	120	4.75	ISO 9660 (1988)	Volume & file structure for information inter- change
			ISO 10149 (1989)	Data interchange on CD-ROM
REWRITABLE OPTICAL READ ONLY	90	3.5	ISO CD-10090	Only one format, intended as re-writeable, actually 86mm, allows read only for software distribution

**Notes:**

- 1) A standard file structure for removable optical media is pending. It will be a logical file structure vice a physical file structure. The standard will be for WORM drives and re-writeable drives, not for CD-ROM. The standard will be an information interchange standard, with logical file structure, comparable to the interchange ability now available with magnetic tape. (Hallam, 1990)
- (2) International Standards Organization (ISO); Draft Information Standard (DIS); Committee Draft (CD)
- (3) Information displayed in this table was obtained from the International Standards Organization (ISO, 1991) and from Mr. K. J. Hallam, Chairman, American National Standards Institute (ANSI), X3B11 Committee (Hallam, 1990; Hallam, 1991)

format any CD-ROM disc drive can read any CD-ROM disc mastered in accordance with the standard, regardless of the manufacturer. Applications may vary, but the physical and logical format of the CD-ROM discs will be uniform.

#### **B. RECORDING AND READING TECHNIQUES FOR OPTICAL MEDIA**

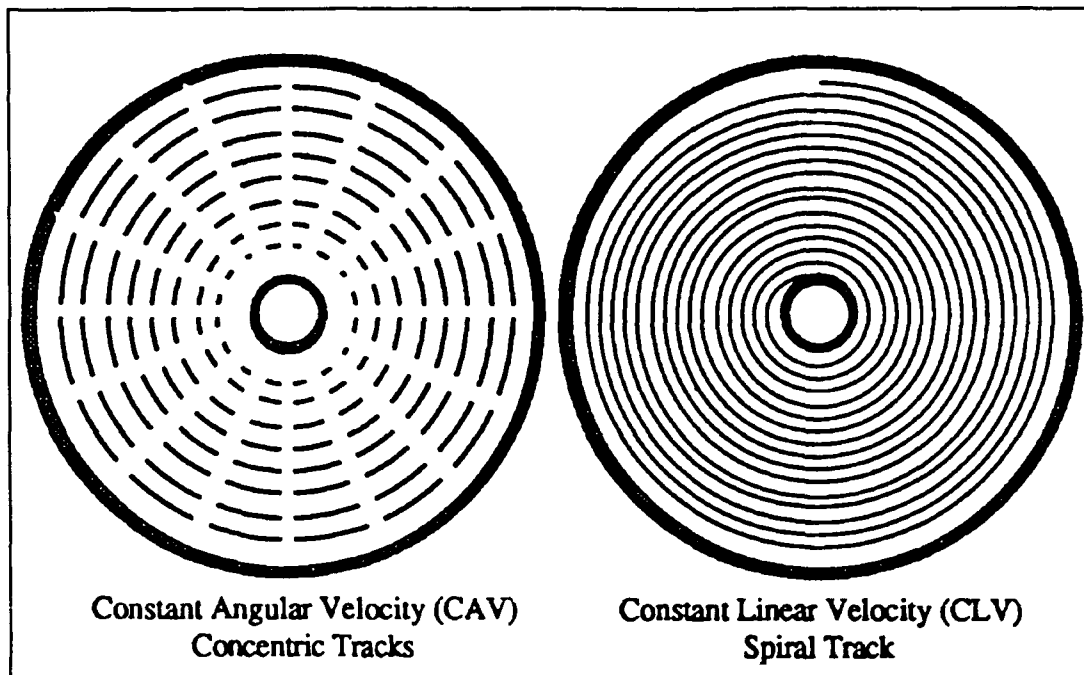
The recording techniques are referred to as ablative, thermal-bubble, or amorphous/crystalline. In the first two techniques, a binary digit is recorded when a small high density laser beam strikes the recording layer of the metal surface of the disc, thus creating a pit, a bubble, or a color change. In the third method a laser sensitive material is altered from a non-reflective to a reflective state.

These state changes can be detected by using a light source in the reading process. Reflective surfaces between two non-reflective surfaces (pits or bubbles) are referred to as lands. A low intensity laser is focused on the track of the disc. Light is diffracted by the pits and is reflected by the lands. The amount of light reflected back into the objective lens is then measured. Modulated signals produced by the combinations of reflected and diffracted light are the representations of the stored information. (Lambert, 1986)

### C. COMPACT DISC READ ONLY MEMORY (CD-ROM)

The rotation technique used by CD-ROM is constant linear velocity (CLV). CLV means that the rotation speed varies according to the location of the disc being accessed. The speed varies from 200 to 500 rpm. The rotational speed accelerates when the inside tracks are being read and slows down when the outside tracks are read.

Figure 1 depicts how the spiral track of a CD-ROM is organized. There are 16,000 tracks per inch on a CD-ROM and the tracks are referenced in minutes, seconds, and sectors. This feature provides massive storage capacity, but also contributes to a relatively slow retrieval time. (Buddine, et al., 1987) The physical addressing scheme of CD-ROM originated from Compact Disc Audio (CD-A). A CD-ROM disc can hold 60 minutes of data. Each minute is divided into 60 seconds. A second of data contains 75 sectors. Therefore a CD-ROM disc contains 270,000 sectors. Each sector contains 2 kilobytes of information, not including the synchronization data, header data, error detection code, unused space, and error correction data. Therefore, the data storage capacity of a CD-ROM is 540 megabytes of information. (Ropiequet, et al., 1987) Table 2 illustrates the allocation of storage space within a CD-ROM sector. Table 3 details the physical format and storage capacity of CD-ROM.



**Figure 1.** Comparison of CAV and CLV formats (Meridian, 1990)

**TABLE 2. STORAGE ALLOCATION OF A CD-ROM SECTOR**

Synchronization data	12 bytes
Header data	4 bytes
User data	2048 bytes
Error detection code	4 bytes
Unused data	8 bytes
Error Correction data	276 bytes
<b>Total</b>	<b>2352 bytes</b>

**TABLE 3. STORAGE CAPACITY OF A CD-ROM DISC**

Minutes per disc	60	
Seconds per minute	60	
Sectors per second	75	
Total number of sectors	270,000	
Total capacity of a sector	2,352	bytes
Usable capacity of a sector	2,048	bytes
Total capacity	635.04	mega bytes
Total user data capacity	552.96	mega bytes

Many authors compare average seek times of CD-ROM to magnetic media. However, comparisons of this nature mask the real advantage of Compact Disc publishing. CD publishing addresses a different environment than magnetic storage. Its purpose is to provide wide distribution of stable information. Information distributed using this medium is not constantly updated, but it is primarily intended for reference purposes, e.g., manuals and other forms of documentation. Conversely, magnetic media are better for information intended to be updated frequently, e.g., on-line, real-time database applications.

CD-ROM publishing has enjoyed a relatively broad distribution. Examples of commercial applications include: "The American Heritage Dictionary", Roget's II Electronic Thesaurus", "Bartlett's Familiar Quotations", "The Chicago Manual of Style (13th Edition)", "The Houghton Mifflin Usage Alert", "The Houghton Mifflin Spelling Verifier and Corrector", "The 1987 World Almanac and Book of Facts", "U.S. Zip Code Directory", and "Business Information Sources" all on one disc. (Bonner, 1990)

Large business organizations have also become heavily involved with CD-ROM applications. For example, Arthur Anderson and Co. publishes all of their reference material for use by the firm's professionals during site visits on CD-ROM, thus allowing easy access to vast quantities of information without transporting large volumes of books. The Ford Motor Company, Agricultural Machines Division publishes all of the information available on their parts and components from the divisions product line on CD-ROM. Mack Trucks Inc., also publishes parts information on their 487,000 custom trucks on CD-ROM. The Army Corps of Engineers Printing and Publication Management Office converted their manuals, specification guidelines, and procedural guides to CD-ROM. The DOD Hazardous Materials Information System is being migrated from microfiche to CD-ROM. (Bonner, 1990)

Recording a CD-ROM requires that procedures be carefully followed. The requirements are outlined below, as recommended by Lind (1987).

1. A concise definition of user requirements, to include data requirements, reporting formats, and dialogue management.
2. Definition of the delivery system, including a detailed description of hardware and software including the equipment manufacturer, operating system, and application system.
3. Data collection via key-board, optical character recognition (OCR), or image scanning. Data collection is very labor intensive, and cost estimates are a critical part of the system design process.
4. Data conversion of machine readable media to a format compatible with index and retrieval software. File structures must match the delivery system. Data may also need to be re-blocked, encrypted, compressed, or edited. Like item 3 above, this function is also labor intensive.
5. Inverted indexes of full text documents are prepared, indexing of key fields, and cross referencing, compression and encryption are preformed as necessary.
6. Software, data, associated indexes, and retrieval structures must be assembled. Directory managers must be constructed, and the disc image must also be determined. In this step, pre-mastering is accomplished. This usually is done by a service bureau. All of the data is transferred to a 1/2" tape. The tape format is verified and error detection and correction codes are calculated.
7. Mastering is the final step in recording a CD-ROM. The tape is converted to analog format for recording. Then a high-powered laser is used to burn data into a glass master. A negative impression is taken in metal and used as a stamp. Replicas are made using multiple polycarbonate discs, which are then coated with a thin layer of metal and coated with protective lacquer.



The outline above demonstrates that producing a CD-ROM requires the same analysis, planning, design, and execution as developing any automated system, and more. Recent advances in the CD-ROM field have enabled mastering in an office environment, versus a sterile environment. This creates a significant cost savings.

#### **D. WRITE-ONCE READ MANY (WORM)**

Until recently, standards have not been universally accepted by WORM manufacturers. However, the apparent lack of a standard for WORM disc drives has not greatly impeded their acceptance in the marketplace. This is illustrated by the fact that several organizations have made significant commitments to the technology.

For example, the United Services Automobile Association has invested more than \$130 million in WORM technology, the Delaware Secretary of State's office converted all of its microfiche to optical disc, the Department of the Army has contracted to migrate its personnel records to optical disc, and the Department of Defense has included WORM drives in its Desk Top III contract.

WORM technology is well suited for document filing applications. Documents may be placed into storage on a WORM disc on an ad-hoc basis using an image scanner.

Records stored in this manner cannot be altered, but they can be updated. Updates are accomplished by appending new documents to the "folders" of existing documents. The new file is "linked" to the old one.

The rotation technique used by WORM disc drives currently on the market is Constant Angular Velocity (CAV). The CAV technique divides the disc into a set of pie shaped sectors, and a series of concentric circles. Figure 1, illustrated the CAV format. This technique is similar to that used in magnetic media. CAV allows tracks and sectors to be directly addressed. CAV allows faster retrieval of data than the CLV technique, but provides a lower storage capacity.

The storage capacities of different WORM discs depend on the diameters of the discs and the formats used. The storage capacity of 300mm (12 inch) discs is approximately 1 gigabyte. The storage capacity of 130mm (5.25 inches) discs varies between 200-400 megabytes, depending on format and manufacturer.

#### **E. OTHER TYPES OF OPTICAL DISC**

There are several other types of optical recording methods including: Compact Disc Interactive (CD-I), Compact Disc Programmable Read Only Memory (CD-ROM), Compact Disc Video (CD-V), Magneto-Optical, and Thermo-magnetic. In this paper we will limit our discussions to CD-ROM and WORM

technologies; the two technologies that are currently best suited for archival applications.

## **IV. OPTICAL DATA-ACQUISITION SYSTEMS**

### **A. OPTICAL SCANNING HARDWARE**

#### **1. Three Types of Scanners**

There are three basic types of hardware configurations for optical scanners: moving paper scanners, flat bed scanners, and electronic digitizing cameras.

##### **a. Moving Paper Scanners**

Moving paper scanners are based on facsimile (commonly referred to as "fax") technology. Documents are conveyed by a transport mechanism past a fixed optical scanning device. These kinds of scanners are less expensive than flat bed scanners or electronic digitizing cameras. Because of their automatic paper feed capability moving paper scanners are a good choice for a mass conversion application or an application where large quantities of documents must be scanned. But like the automatic feed mechanisms in popular office copy machines, problems can occur with the document transport mechanism (paper jams, etc.). (Waegemann, 1989)

##### **b. Flat Bed Scanners**

Flat bed scanners are based on copy machine technology. Documents are placed on a glass platen and the optical scanning device is mounted on a carriage that is passed under the document. Flat bed scanners are more

expensive than moving paper scanners, primarily because the carriage required to transport the optical scanning device drives the cost up. However, flat bed scanners are less expensive than electronic digitizing cameras. Flat bed scanners are the best choice for work-station peripherals or desk top publishing applications where the volume of documents to be scanned is not excessive. Flat bed scanners perform well for applications such as: scanning photographs, oversized objects, pages from books, or where precise positioning is important. (Waegemann, 1989)

### **c. Electronic Digitizing Cameras**

Electronic digitizing cameras are based on camera technology. They utilize a camera that has replaced film with an optical scanning device. Documents are placed on an image plane and a stepper motor or a servo-drive system positions the camera. This procedure occurs under the control of the host central processing unit (CPU). Electronic digitizing cameras look like microfilm cameras and were actually the first digital scanners. This scanner is by far the most expensive type. However, an electronic digitizing camera is quite flexible and can be used to scan oversized items that can not be scanned using moving paper scanners or flat bed scanners. (Waegemann, 1989)

## **2. A Description of the Scanning Process**

The scanning process has several steps that are basically the same for all three types of scanners. The

primary difference between scanning technologies is the method of document transport, as discussed above. This process involves two major components: a low frequency light source and a charge-coupled device (CCD). The CCD is an integrated circuit that converts light into digital information.

**a. The Charge Coupled Device (CCD)**

The CCD is a photo converter that is used in most scanning machines. It is a light-sensitive semiconductor that produces electrical charges based on the light incident on its surface. In this process an analog image is converted to a digital representation of that image and is referred to as a raster image. The photocells on the surface of the CCD convert the optical signal into an electrical signal. The voltage of the signal is proportional to the intensity of the optical signal. The white areas of the original image reflect more light and therefore generate greater voltage. (Stanton, et al., 1986)

**b. The Light Source**

A low frequency light source illuminates a strip of the document with each movement of the document in a moving paper scanner or of the carriage in a flat bed scanner. The light is reflected by the light areas and is absorbed by the dark areas of the paper. Mirrors pass the light reflected from the document to a lens. The lens focuses the reflected light onto a photodiode array, or a

charge coupled device (CCD). The CCD transforms the optical signals to digital signals.

(1) The Vertical Scan. The vertical scan process occurs as the light source moves through the original document line by line. The distance between the lines to be scanned depends on the resolution setting (e.g., 151 scan lines per inch). As the light source pauses on each vertical scan line the horizontal scan takes place.

(2) The Horizontal Scan. During the horizontal scan information from the illuminated strip is "read" and converted to a digital format. The strip that is illuminated by the vertical scan is divided into sections. The size of each section is determined by the resolution settings in pixels per inch (ppi). (Taylor, 1989)

## **B. Image Scanning**

### **1. The Two Key Characteristics of Image Scanning**

The two key characteristics in image scanning are resolution and greyscale. Unlike coded formats such as ASCII, image scanning stores graphic and text images as two dimensional bit maps; the information is not directly addressable. The key characteristics of image scanning are discussed in the following sections.

#### **a. The Resolution of a Scanned Image**

Resolution in image scanning refers to the density of the dot-matrix representation of the image and is

measured in dots or picture elements (pels) per square inch. The greater the resolution, the finer the detail. A resolution of 75 - 100 pels per inch (ppi) is of a good quality, but details are hard to detect. A resolution of 200 ppi has a quality equal to, or greater than, most original documents. Resolutions of 300 ppi and above have a quality greater than most originals. In these comparisons, the term original document refers to an original page produced by a typewriter. (Taylor, 1989)

#### **b. The Greyscale of a Scanned Image**

Greyscale refers to the number of shades of grey to be used in representing an image. Greyscales are represented by picture elements (pixels). Pixels represent more information than the previously introduced pel, including information such as color, brightness, and intensity.

Greyscales are required to represent the continuous tones of originals such as photographs. A greyscale has several components, including: thresholding, halftoning, windowing, and compression. These components are described below.

**(1) Thresholding.** Thresholding is a technique used to convert images into binary descriptions. A particular shade of grey is selected as the system's threshold. Shades of grey lighter than the threshold are



represented as "zeros" while shades of grey that are darker than the threshold are represented as "ones".

(2) Halftoning. In halftoning, greyscale information is processed to create a higher level pattern of dots in certain areas to produce shades of grey. Basically, the more dots that are in an area, the darker the area appears. This technique is used for high-quality images or photographs. Pictures in newspapers are examples of halftones. The technique is also used in radiographs (x-rays).

(3) Windowing. In windowing, the first scan of a document uses thresholding to scan the graphics. A window is then placed around the graphic image and halftoning is used in the second scan to optimize the image.

(4) Compression. An enormous volume of information is generated in the process of scanning images and a large amount of storage is required to store this information. Electronic imaging would be infeasible without compression. By employing mathematical algorithms, the white space in images can be represented in a more concise form. Using compression one square inch of white space can be described by a few bits vice thousands. Compression algorithms were first developed for facsimile transmissions, and subsequently were standardized. They are described in the International Telegraph and Telephone Consultative

Committee (CCITT) group 3 and 4 standards. (Waegemann, 1989)

## **2. The Practical Limitations of Image Scanning**

Practical limitations must be considered in designing image scanning systems. Current technology can scan resolutions up to 2000 ppi, and describe up to 256 grey tones. Table 4 lists the number of bits required to store various levels of greyscale. The calculations for computing the requirements for storing images are listed in Table 5. For example, the storage required for an image with dimensions of 8.5" x 11" and a resolution of 2000 ppi and 256 greyscales would be 23.936 billion bits. That would, indeed, be an expensive page to store. An eight layer greyscale at 200 dpi would require 33.66 million bits of storage. Most laser printers can only reproduce greyscales of 64 grey tones (6 bits per pixel). High resolution printers and display devices are available, however, their costs may be prohibitive.

## **C. Optical Character Recognition (OCR)**

### **1. Two Types of Optical Character Recognition**

There are two types of optical character recognition: matrix matching and topographical analysis.

**TABLE 4. STORAGE REQUIREMENTS FOR GREYSCALE IMAGES**

Levels of greyscale	Bits per pixel
256	8
64	6
8	3

**TABLE 5. STORAGE REQUIREMENTS OF A RASTER IMAGE**

Parameter	Basis
Pixels Per Inch (ppi)	resolution
Bits Per Pixel (bpp)	greyscale
Base of the image (B)	inches
Height of the image (H)	inches
Storage requirements (S)	bits
$S = B \times H \times (ppi \times bpp)^2$	

The unique features, capabilities, and disadvantages of each type of character recognition are discussed below.

**a. Matrix Matching**

Matrix matching is a form of OCR in which a scanned character is compared with a set of templates for each font that the system can "read". Multi-font matrix matching systems require increased memory capacity to store the fonts supported and to perform the comparative analyses. This method of OCR is sensitive to subtle differences in character shapes, however it is relatively insensitive to broken characters.

Matrix matching technology is comparatively fast and has a high degree of accuracy. The accuracy is reported to be as high as 99.9 percent, or two errors per page. Matrix matching is able to handle poor quality originals including third generation photocopies. Its disadvantages include its lack of capability to recognize most typeset characters, limiting it to the most common typewriter and printer fonts. (Mueller, 1988)

**b. Topographical Analysis**

Topographical analysis is also referred to as feature extraction. In this method important features of a characters image are used to determine what character is being represented. Features are defined as vertical and horizontal strokes, line endings, closed and open curves, slanted strokes, and intersections of strokes. This method

is relatively insensitive to slight variations in the shape and sizes of characters, and less memory is required for font libraries. However, a disadvantage of topographical analysis is sensitivity to broken characters.

An advantage of topographical analysis is that it can be used in intelligent character recognition software. Intelligent character recognition (ICR) is a form of artificial intelligence. The system can "learn" via operator assistance. Operators can intervene to identify characters that the system can't identify. ICR is an improvement over OCR and is touted as the key to future success of conversion scanning.

#### **D. Performance Characteristics**

##### **1. The Effects of On-Board Processing Power**

Scanners often have their own on-board processing power and memory. These features can be located in the scanner or on the interface card. Scanners can also rely on the host system for processing power and memory. The main advantages of having the capability on-board the scanner are device independence and an ability to work in the background. This means while scanning and image processing tasks are being processed, the host computer is available for other jobs, such as word processing. These features allow higher performance of the scanner and they free the host computer for other applications.

## **2. Accuracy of Scanning**

Using the matrix matching method of OCR the accuracy rate is reported to be about two to three errors per page. The accuracy of topographical analysis depends on the set of algorithms used to describe each character and the particular tool's ability to "learn". Most scanners equipped with topographical analysis technology can be trained to unique type faces. Top-of-the-line tools employ artificial intelligence, and the tool's ability to interpret new type faces depends on the number and the capability of the expert modules employed in the system. (Mueller, 1988)

Accuracy is a matter of resolution in image scanning. A resolution of 200 ppi will produce images that are equal to, or greater than, original document resolution. The facsimile standard, conforming to the CCITT groups 3 and 4 standards for compression algorithms, is 200 ppi. (Mueller, 1988)

## **V. DOCUMENT RETRIEVAL SYSTEMS**

The retrieval system is the most critical link in any optical based storage system. If the documents are not available when needed, the system is of no value. If documents are stored in their original paper form, no matter how poorly they are filed, then a researcher can do a laborious visual search of the files and still be able to locate a specific document. However, if documents are placed on a disc, then a manual procedure is no longer possible. The documents will only be accessible via the file structure used to place the documents on the disc. It is therefore imperative that a high-quality storage and retrieval system be used to provide quick, effective retrieval capabilities and prevent the loss of any documents. The retrieval system embodies the user interface for the system and will influence the acceptance of the system by the users. For the reasons cited above, retrieval software should be carefully evaluated with consideration given to its potential impact on the entire system.

### **A. DOCUMENT RETRIEVAL DEFINED**

Searching a document base for documents containing information is quite different from querying a database. Document storage and retrieval systems provide access to

documents just as a database management system provides access to data but there are significant differences in how these tasks are done. Blair and Maron, in their 1984 study, pointed out four primary distinctions between document and data retrieval (Blair and Maron, 1985). These four distinctions are discussed below.

#### **1. Document Retrieval is Less Direct**

Document retrieval systems answer inquiries less directly than data retrieval systems do. Document retrieval relies on the assumption that groups of words can be used to approximate meaning. While a data retrieval system would respond to a query for the population of the United States in the 1990 census with the number, 249,632,692, a document retrieval system would provide a group of documents containing the search words "population" and "United States" and "1990". The user could then browse through the documents to determine which of them suited his purpose.

#### **2. Document Retrieval is Probabilistic**

Document retrieval is probabilistic and will not necessarily return documents of value. While data retrieval will either return the queried value or not, document retrieval may return a group of one or more documents that may or may not contain documents which pertain to the query. It remains for the user to decide if the retrieved documents fit his purposes.



### **3. Utility versus Correctness**

Success in document retrieval is measured in terms of usefulness rather than in terms of correctness. For this reason it is far more difficult to measure success in a document retrieval system than in a data retrieval system. A data retrieval system either returns the correct answer to the query or it does not. A document retrieval may return documents that have varying degrees of usefulness.

### **4. Retrieval Time is User Dependent**

In document retrieval, the user's time, not the machine's response time, determines retrieval speed. In data retrieval there is a one-to-one correlation between query and response - this is not the case with document retrieval. The document retrieval process is interactive and iterative with the user evaluating the system's responses and refining his queries. Therefore, it is not the fastest system response time that determines retrieval speed, it is the time required to recover the desired information. A slower but more flexible retrieval system that gives the user an opportunity to narrow or broaden searches as he desires could prove to be the faster means of retrieving information. This factor is of particular significance for CD-ROM since its major disadvantage is slow access speed. If a CD-ROM retrieval mechanism is particularly effective, it can outperform systems based on media with faster machine response times.

## **B. IMAGE-BASED VERSUS TEXT-BASED STORAGE**

Information in a document retrieval system can be stored in image- or text-based format. The choice of format will determine the manner in which the information can be retrieved. Image-based document storage consists of a database where records include several key fields and one very large data field consisting of an image of the document. Such a database is highly structured and permits access only by selected key fields. In contrast to image-based document storage, text-based information storage permits indexing of each word in the document retrieval system. This feature provides increased flexibility and functionality over the previously discussed method with regard to accessing information.

## **C. IMAGE-BASED SYSTEMS**

Image-based systems contain digital "pictures" of the pages of a document. They are particularly good at maintaining the original format of documents and have the advantage of being relatively inexpensive to convert from paper or microfiche. Conversion from microfiche to digital images costs from 17 to 30 cents per page depending on volume (Caldwell, 1991). However, even with compression, image-based systems require up to 25 times more storage space than text-based systems, and large image sizes can cause lengthy transmission delays.

A 300 by 300 pixel per inch letter-sized, uncompressed image would take over 15 minutes to transmit at 9600 baud, and over 8 seconds to transmit at 1 megabit per second. Another disadvantage of an image-based system is its dependence on expensive manual indexing. Each document can cost up to 25 cents to index (Rothchild Consultants, 1989).

Since documents in an image-based system can only be accessed via the terms by which they are indexed, the level of skill and detail used for indexing is crucial. If the indexing is done poorly, or if the terms used for indexing become dated, the information contained in the documents will be inaccessible. Many image-based document systems exist today, for example, the Library of Congress and the Naval Research Laboratory, each of which have millions of images stored. Even though these systems provide protection of the original documents, and provide improved access times over paper documents, they still don't have a method for automated searching of the contents of the documents.

#### **D. TEXT-BASED SYSTEMS**

Text-based systems can unlock the information contained in a document base. A text-based system can be automatically indexed using software that produces an inverted index or concordance. This index lists each word in a document and the location of each instance of every word. While such indexes are large and may typically occupy

as much as 35 percent of the size of the original information, they can be automatically generated and they provide quick access to the content of the documents (Naval Publications and Printing Service, 1990). Even with the added space requirement for an inverted index, an ASCII coded, text-based storage system is very compact. A standard letter-sized page will only contain about 2000 bytes compared with the 50,000 bytes of the compressed image of the same page.

One major obstacle to achieving a text-based storage system is the relatively high cost of converting paper or microform to a text-based system. The state of the art in Optical Character Recognition (OCR) still requires significant expensive manual quality assurance. Conversion costs can run from \$2.00 to \$4.50 per page depending on the volume of documents to be converted (Rothchild Consultants, 1989).

The software must be able to provide a 96 percent accuracy in conversion to be economical when compared with re-keying. With poor quality original documents it may be less expensive to re-key the documents than to use OCR. The decision-maker must decide if the advantages to be gained from having information in full-text retrieval format outweigh the costs of conversion. (Anamet Laboratories, 1988)

## **E. THREE TYPES OF ELECTRONIC DOCUMENT RETRIEVAL SYSTEMS**

Electronic document retrieval can be divided into three classes: database document retrieval systems, full-text retrieval systems, and hybrid systems. The nature of the data will impact the type of retrieval system chosen. Highly structured data that can be grouped into fields are suitable for database retrieval while free-form text can be retrieved with any of the three types but lends itself better to full-text retrieval or hybrid. The advantages and disadvantages of each system are discussed below.

### **1. Database Retrieval**

Database document retrieval employs indexes based on the fields present in, or added to, the database. Image-based document management systems employ database retrieval techniques. Key word indexes of the fields in the database provide extremely quick access to the data since a search of one field can be executed far faster than a search of the entire database. Fielded data also allows for range searches on numerical or date fields. For example, without specific numeric value fields it would be impossible to retrieve all reports in the database less than six months old.

A user needs to be familiar with the terminology used to index the fields of interest, when retrieving information using fielded data because the document is only retrievable by that term. Because most documents do not

have a distinct fielded structure, the fields must be manually designated. This introduces a subjectivity into the indexing. The indexer must make decisions regarding the terms which can be used to retrieve a document and in so doing he determines the usefulness of the document base. In addition to being a very challenging task, indexing is a very labor intensive process and can be quite expensive.

## **2. Full-text Retrieval**

Free-text documents are best suited to indexing and retrieval through full-text retrieval. Full-text retrieval does not tie the user to the limited set of key-words and fields generated by an indexer. Automatically generated inverted indexes containing all the significant words in the database provide direct access to the content of documents. Words not deemed to be significant due to a high frequency of occurrence - stopwords - are omitted from inverted indexes in order to reduce the index size.

Searching for relevant documents based on the occurrence of specific words in those documents is a process that is not guaranteed to produce retrieval sets that contain the desired information. Synonyms, euphemisms, and even misspellings complicate the already significant problems of precision (obtaining only the information desired), and recall (obtaining all the information desired). Since the process of full-text retrieval may, or may not, return relevant documents, systems which employ

this method must provide additional features and flexibility to the user to deal with this uncertainty.

### **3. Hybrid**

Advantages of both types of searching can be obtained by combining the two methods. Many commercial products are doing this today. For example, the full-text of each document may be placed in an inverted index and eight to ten additional fields may be indexed for each document. A user can then either search the text or select a field search which will only look at a specified field. This combination is more expensive to produce than a single method but it provides the user the most flexibility and functionality.

## **F. RETRIEVAL SOFTWARE FEATURES**

The goal in document retrieval is to extract documents from a document storage system that contain information that is relevant to a user's search. Relevance is a subjective term that refers to how well a document relates to a user's needs.

### **1. Full-text Retrieval Features**

Full-text search software can provide a wide range of capabilities. These capabilities have a great impact on the utility of the retrieval software and should be investigated carefully before making a selection. The most important features are discussed below.

#### **a. Phrase Searching**

Any full-text search system must perform phrase searches. The user enters the word or words to be searched and the retrieval software returns a number of documents containing each word and the total number of documents containing any of the words. The user can either view all of the documents selected, or he may refine his query further if the set is too large.

#### **b. Proximity Searching**

The presence of the words "optical" and "storage" in a document does not guarantee that a document containing those words will be relevant to a search for information on optical storage. However, the presence of the two words "optical" and "storage" in sequence, or within three words of each other does increase the probability of the retrieved document being relevant. It is important, therefore, that the retrieval software contain the ability to designate proximity of the search terms. This requires that the index include the additional information of a word's distance from known delimiters such as sentence, paragraph, and document boundaries.

#### **c. Boolean Searching**

Boolean searching involves the use of the AND, OR, and, NOT operators to construct searches. The use of AND between two terms restricts the search by excluding documents which do not contain both terms, while the use of



OR widens the search by including documents which contain either word. The NOT operator provides flexibility in designing queries and also serves to restrict searches.

#### **d. Back Referencing**

The use of boolean searching in an iterative manner to further refine or expand a search is a very useful function. Back-referencing is used to combine an existing retrieval set with a boolean search and to obtain a modified retrieval set.

#### **e. Cross-Referencing**

Cross-referencing is the ability to browse through related documents either by using manually inserted links which take the user to documents containing related information or by executing another query. The ability to move in a non-linear fashion throughout the document base is one characteristic of a hypertext system and is useful in gaining general knowledge of a subject.

#### **f. Query Expansion**

Variations in the spelling or form of a word can prevent a user from retrieving relevant documents. The retrieval system should have the capability to expand a search to include plurals as well as other forms of root words. This capability could also allow for some misspellings by extracting the root word and appending the properly spelled prefix or suffix for the user. The users'

needs for speed and functionality must be considered when making the decision to add this feature.

#### **g. Thesaurus**

Another type of query expansion involves the use of a thesaurus. A query can be expanded to include synonyms, abbreviations, and technical jargon relating to the query term. The expansion process simply uses the Boolean OR operator to widen the search for the synonyms.

#### **h. Browsing**

An effective marriage of searching and browsing is essential to an effective document retrieval system. Searching, especially full-text searches on computer-generated inverted indexes, will get the user to a retrieval set of documents, many of which contain relevant information. From there, browsing will let him fine tune his research and focus on those documents that have true relevance to his subject. The ability to browse documents on-line and to decide quickly whether or not a document is relevant provides a researcher a most effective tool.

### **2. Database Retrieval Features**

#### **a. Field Searching**

This feature provides a quick access to documents with a fielded structure. The software need only search the specified field's index for the search terms and can therefore perform a very rapid search.

### **b. Range Searching**

Searching for a range of values is only possible if the data has been entered into fields and the fields indexed accordingly. Numerical and date data are best stored in fields so they may be retrieved in range searches.

### **G. SELECTION CRITERIA**

The functionality discussed above as well as the costs for acquiring and licensing the software must be considered in the selection of a retrieval software package. Packages providing full-text and database retrieval capabilities are available from \$995 to \$15,000 or more for custom requirements and vary widely in the capabilities provided. Most of these systems are capable of handling combinations of text and images which is essential if entire documents are to be stored. Ease of learning and use should be evaluated since these factors could be critical to the acceptance of the system by end-users. Appendix A contains a checklist to be used for evaluating retrieval software packages.

### **H. IMPORTANCE OF RETRIEVAL SYSTEMS**

The value of a document retrieval system lies in its ability to retrieve information when needed. The functionality and quality of the retrieval system, therefore, will determine the value of the system. All the costs of conversion and storage will be for nought if an

ill-suited retrieval system is put into place. Any decision to establish a document storage and retrieval system should begin with consideration of the retrieval system and how it will affect other aspects of the system. Sufficient resources should be devoted to both evaluating and acquiring the appropriate software for each document retrieval application, given the critical nature of retrieval systems.

## **VI. TECHNOLOGY FOR MIGRATING IMAGES TO OPTICAL DISC BASED SYSTEMS**

### **A. THE NEED FOR IMAGE MIGRATION**

Many Federal government agencies are in the process of learning how to migrate their information bases to optical disk storage devices. The Library of Congress, the National Archives and Records Administration, the U.S. House of Representatives, and the Department of Defense are examples of large organizations that currently have optical disk projects in progress.

The majority of these initiatives are focused on the acquisition of information contained on paper and in computer systems. There remains, however, a need to migrate information currently stored on microfiche to an environment where it may be categorized, described, and quickly retrieved. Examples of these kinds of applications include military medical and personnel records stored on microfiche.

The degree of flexibility in manipulating information stored on microfiche is severely limited. Instant availability of images, multiple user access, and relational search potential are not possible in microfiche-based systems. These additional capabilities are available in the media of optical disk, and they greatly expand the range of potential applications.

The technology for transferring images based in microfiche to optical disk systems has existed for a number of years. Several organizations either have already accomplished this type of migration or are in the planning process. Nevertheless, literature describing the technology, and the methodology used in evaluating it, is not readily available. Therefore this chapter will provide a description of the technology used to capture microform images and transfer them to optical media.

#### **B. EARLY RESEARCH INTO MICROFORM SCANNING**

The Federal Government's continuing interest in microfiche scanning is demonstrated by several research reports developed during the 1970s. A report issued for the U.S. Air Force by Singer-General Precision, Inc. in 1971 focused on the problem of updating microfiche.

The requirement to update the information on the microfiche posed numerous problems. The primary problem being the high volume of microfiche retained by the Air Force. Microfiche are exposed diazo film and can not be updated incrementally. If a frame must be updated the entire fiche must be reproduced. This limitation of microfiche (since solved by AB Dick updatable microfiche, and jacketed microfilm) presented the Air Force with the problem of having to retain large volumes of original documents to enable them to reproduce the microfiche if an

image needed to be updated. Although microfiche can be copied, and updated, the image is degraded in the copying process. Microform, under the best circumstances, can only be copied 5-10 times. The image quality of each copy is lower than the preceding copy and the legibility degrades in each generation. (Hayes, et al., 1971)

The alternative analyzed in the Air Force study focused on the development of a human-readable and machine-readable microfiche (HRMR). The HRMR microform stores a digital representation of the image on the microfiche itself. This allows duplication of the images without risking their degradation or creating a need to retain the original documents. (Hayes, et al., 1971)

In a report issued by the Naval Undersea Center in 1975, the feasibility of a microfacsimile system was analyzed. The emphasis of this study was the timely and efficient dissemination of Naval personnel records stored on microfiche at the Naval Bureau of Personnel in Washington, DC. This was to be accomplished by scanning microfiche personnel records and transmitting them using facsimile technology. (Endicott, et al., 1975)

Another report written in 1976 by EPSCO Labs for the U.S. Air Force described yet another use for microfiche scanners. This study addressed the feasibility of scanning microfiche and storing them in a digital format. The digitized microfiche were to be stored in a buffer partition

belonging to each end-user on a mainframe computer. Then the end-user could display the "digitized microfiche-reports" on their Tektronics 4041 display terminals. (Botticelli, et al., 1976)

A number of reasons for the design were explained in the preceding paragraph. The primary reason was to expedite the dissemination of microfiche reports. This system was designed to provide very fast access to those images that had been pre-loaded into users' partitions. Another reason for the system design described above was the limitation of the technology available at the time.

Disk storage was expensive in 1976, compared to the cost versus capacity ratios that can be achieved today by using optical disk technology. Storage was limited because of the expenses involved. It was more economical to store the reports on microfiche. Large volumes of storage are relatively inexpensive today with the advent of optical disk storage technology.

### **C. MICROGRAPHICS TO OPTICAL CONVERSIONS IN PROGRESS**

There are numerous on-going initiatives within the Federal Government, and in other organizations, to convert microfiche holdings to a digital format stored on optical disk. Examples of organizations reporting these initiatives are recounted below, but this by no means is a comprehensive listing.



The Library of Congress began an optical disc pilot project as early as 1983. A prototype microfiche scanner was included in this project as reported by Manns and Swora, 1986. In a discussion with Mr. Manns, it was determined that the results of the Library of Congress' attempts to digitize microfiche were successful. High demand items from the retrospective collection were converted to a raster format, and Manns (1990) reported that the scanner did a very good job. The LOC has plans to convert the existing microfiche collection to a digital format.

The Delaware Secretary of State's office recently converted their microfiche to optical disk, as reported by Butler, 1990. In this conversion, due to stringent quality control standards, the error rate was less than one percent.

The U.S Army has reported a very ambitious project to convert all of their personnel records to optical disc. Table 6 details the large number of these records that are currently stored on microfiche. Lingvai (1991) reported that the contract for converting the Army personnel records has been awarded, and that conversion is in progress.

The U.S. Navy has initiated a project related to the migration of microfiche to optical disc. The Engineering Data Management Information and Control System (EDMICS), has been reported to be the largest engineering document

**TABLE 6. IMAGES TO BE CONVERTED IN THE U. S. ARMY'S PERMS PROJECT**

Location	Records	Type	Media
MSD	107,377	OMPF	MICROFICHE
EREC	652,198	OMPF	MICROFICHE/PAPER/ODI
GUARDPERCEN	52,811	OMPF	MICROFICHE/PAPER
ARPERCEN	589,561	OMPF	MICROFICHE/PAPER
	304,901	MPRJ	PAPER
	710,768	MED/SPC	PAPER
	1,641,000	RETIRED	MICROFICHE/PAPER

**Notes:**

Enlisted Records and Evaluation Center (EREC)  
 National Guard Personnel Center (GUARDPERCEN)  
 Army Reserve Personnel Center (ARPERCEN)  
 Official Military Personnel File (OMPF)  
 Member Personnel Record Jacket (MPRJ)  
 Medical/Special (MED/SPC)

This information is based on a presentation by Lingvai, (1990)

management project in the United States. Engineering documents have been traditionally stored on aperture cards (a frame of 35mm film placed in a tabulating card).

In early testing the contractor that won the contract demonstrated the ability to scan 900 aperture cards per hour (four per second). (Kaebnick, 1990) The project manager reported that a review of the Louisville test site is scheduled for April 1991, and if approved the project will expand to 43 Navy sites and four commercial shipyards (Kyle, 1991).

#### **D. READER-PRINTERS AND READER-SCANNERS**

There is a significant distinction between reader-printers and reader-scanners. A reader-printer is a device that uses optics to produce an analog representation of a microform image on dry silver paper (the latest models that are actually reader-scanners print to copier paper.) A reader-scanner has been described as a "new type of reader-printer" that converts microform to a raster image (Burnacz, 1990).

However, that definition is incomplete. A reader-scanner can perform the role of a reader-printer, but why stop at that point? A reader-scanner produces a raster image of a microform image. Once a bit image is in the users' control, potential uses of it are only limited by the users' imaginations. The image can be transmitted by

telecommunication, stored on optical disc for future use, cropped or windowed, converted to ASCII text, displayed on a computer terminal, or printed on a digital-laser printer.

#### **E. MICROFORM IMAGE SCANNERS**

There are numerous microform scanners available on the market. During the 35th Annual Conference of the Association of Records Managers and Administrators, held in San Francisco between 5-8 November, 1990, numerous major corporate vendors displayed their microform scanners. Most were marketed under the name reader-scanner, while others used the terminology, microform digitizing and image scanning.

The prime difference in the terminology is the intent of the usage of the equipment, not the technology. Reader-scanners, as described above, are intended to produce raster images for the purpose of printing, or in some cases for facsimile transmission. Microform digitizers, or microform image scanners, are intended to transmit the raster image to a computer system.

A microform image scanner is not a great deal different from a paper scanner. The primary difference is that a microform scanner uses optics to magnify the images, which are then scanned with a charged coupled device (CCD) array. Another difference, in microfiche scanning, is the use of an x-y transport to position the microfiche. Figure 2 presents

a schematic drawing of the components of a microfiche scanner. (Burrus, 1990; Douglass, 1990)

The operation of a microfiche and flat bed paper scanners are similar because the operators of both scanners place the microfiche, or paper, document on a glass platen. The difference is that once placed on the platen, the microfiche scanner will now position the fiche, and scan all 98 frames of a 24x microfiche - at a rate of 33 frames per minute; while the paper scanner does no positioning and scans only a single page at a time. (Burrus, 1990).

The process of scanning microfilm is quicker and easier than scanning microfiche. This is simply because roll microfilm is continuous. The microfilm is placed on an output spindle and a take-up reel, much like a microfilm reader. Microfilm is then passed by an optical device that magnifies the images, and is continuously scanned by a high resolution linear array camera. (Mekel, 1989)

Microform scanners, like paper scanners, can produce image resolution of between 300 to 400 dots per inch (dpi). This produces a large raster image. An 300 dpi image with an aspect ratio of 8.5" x 11" creates a frame size of 2550 x 3300 pels, requiring 8,415,000 bits of storage. If we had selected 400 dpi, then we would have produced 14,960,000 bits of storage.

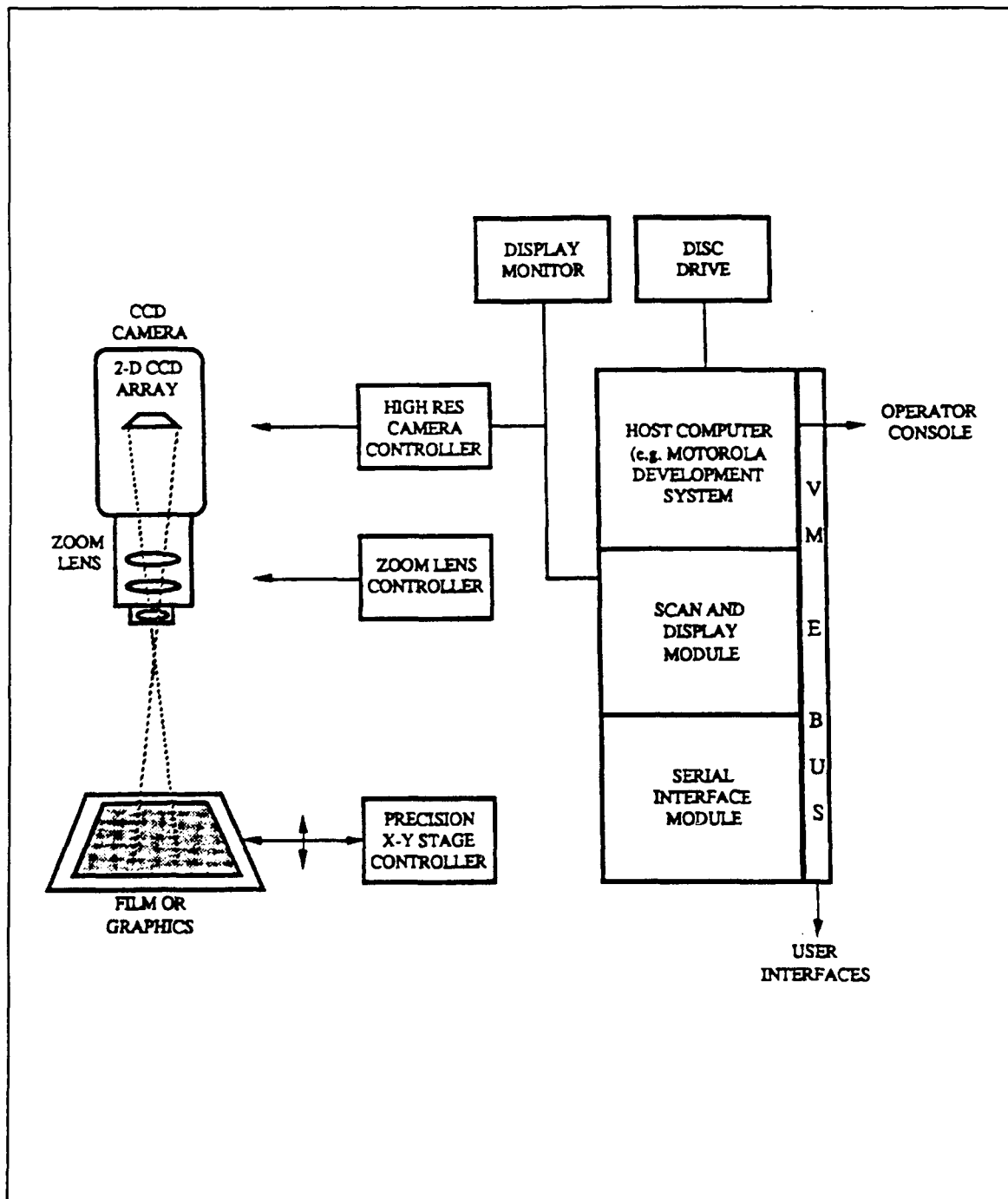


Figure 2 Schematic diagram of a microfiche scanner (Douglass, 1990)

Data compression is important to enable manageable handling and storage of this information. Mekel Engineering Inc., reports the following storage requirements for one image digitized at 200 dpi. Using a compression ratio of 12:1, the image cited requires 20 kilobytes per image. Fifty images require one megabyte, and 1000 images require 20 megabytes. (Mekel, 1989) Based on these figures, one 24x microfiche, consisting of 98 frames, will require at least two megabytes of data storage.

#### **F. THE FEASIBILITY OF MICROFORM DIGITIZATION**

Microform can be successfully converted to a raster image, and there are important initiatives in progress to accomplish these ends. However, the storage requirements of these images are greater than can be reasonably accommodated by magnetic disk. To meet these demands the higher capacity of optical storage is required.

The technological possibility of an endeavor is only part of the feasibility analysis. Other considerations are it's cost, and the point at which the costs incurred by making the change are outweighed by it's benefits. In other words - what is the value of the information?

If highly paid staff, such as scientists, engineers, doctors, attorneys, and others require the information on a regular and recurring basis, then the cost of conversion may be justified. It may also be worth the effort and expense

if the information is critical to security, health care, or safety. An organization needs to rigorously investigate the costs associated with these benefits. A thorough analysis of the value of the information to the organization will help to avoid the pit-falls of racing ahead blindly and embracing the latest technology.



**VII. ANALYSIS OF THE REQUIREMENTS  
FOR MIGRATING UNCLASSIFIED TECHNICAL REPORTS  
FROM MICROFICHE TO OPTICAL DISC  
IN THE KNOX LIBRARY, RESEARCH REPORTS DIVISION**

**A. METHODOLOGY**

The scope of this chapter is to identify the criteria for migrating a microform based information system to an optical storage and retrieval system. A case study was selected as the methodology for this investigation.

The authors believed that a large microform information base, fairly representative of a typical government information system, was essential for the study. The Dean of Computer and Information Services at the Naval Postgraduate School suggested the Knox Library as a good source for the type of information base desired. In a subsequent meeting with the Director of the Knox Library, the Defense Technical Information Center (DTIC), Technical Reports (TR) information base held by the library's Research Reports Division (RRD) was suggested as a suitable subject for this case study.

**B. REQUIREMENTS ANALYSIS**

The mission of the Naval Postgraduate School is "to conduct and direct the education of commissioned officers and to provide such other technical instruction as may be

prescribed to meet the requirements of the Naval service; and in support of the foregoing, to foster and encourage a program of research in order to sustain academic excellence" (Naval Postgraduate School, 1990). The Research Reports Division (RRD) of the Knox Library supports this mission by assisting professors, staff, and students in accessing a wide variety of government research reports.

The scope of our study will be limited to one source of government research reports, the Defense Technical Information Center (DTIC). The function of DTIC has been explained by Jones (1990).

The Defense Technical Information Center is responsible for collection and dissemination of scientific and technical information for DoD activities and their contractors.

This information source is the focus of our study for a number of reasons. First, it is an important and broad based source of technical and scientific information. Second, it is a well defined source of information and reports. Third, the DTIC technical report database is primarily available only through the media of paper, microfiche, microfilm, on-line, and tape products. Fourth, this database is under-utilized by the faculty and students of the Naval Postgraduate School. Table 7 presents the frequency data supporting this conclusion. The authors suspect that the primary reason for underutilization is due to the database storage medium, i.e., microfiche.

TABLE 7. KNOX LIBRARY RESEARCH REPORTS DIVISION MONTHLY TRANSACTIONS

1990 MONTH:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	TOTAL	AVERAGE	DAILY	AVG
FICHE RECEIVED:																
UNCLASS UNLIMITED	1066	1014	1295	1809	1426	503	1498	1225	1275	635	1022	646	13414	1118		56
UNCLASS LIMITED	727	953	854	1314	538	309	1705	10 <sup>1</sup> / <sub>4</sub>	6	394	520	417	8821	735		37
CLASSIFIED	0	173	44	178	109	92	83	219	67	76	152	99	1292	108		5
TOTAL	1793	2140	2193	3301	2073	904	3286	2528	1348	1105	1694	1162	23527	1961		98
FICHE COPIED (CLASS)	904	1434	1050	1379	3118	764	872	2350	549	2018	2223	443	17104	1425		71
FICHE CIRCULATED	128	239	228	506	532	157	272	555	193	508	405	122	3845	320		16
FICHE VIEWERS CIRCULATED	6	4	4	10	10	2	2	9	4	12	5	1	69	6		<1
TECH REPORTS CIRCULATED	765	498	604	1151	928	365	832	998	417	1323	764	417	9062	755		38
COMPUTER SEARCHES	488	341	258	577	429	144	521	441	221	473	296	166	4355	363		18
PATRONS SERVED	162	118	85	199	144	58	180	162	70	138	109	48	1473	123		6
DROLS USAGE																
MINS ACTIVE	2306	8171	9583	8435	8723	7128	6934	10823	7409	9637	8614	6856	94619	7885		394
MINS PROCESSING TIME	294	770	597	1212	920	325	1067	1038	541	1187	650	354	8955	746		37
TR SEARCHES	377	797	675	1175	845	360	923	902	574	991	688	438	8745	729		36
OTHER SEARCHES	1	0	2	21	4	1	6	8	2	9	12	4	70	6		<1
TR DOCUMENTS ORDERED	6	23	8	35	58	12	9	20	9	31	8	4	223	19		1

## NOTES:

HARD COPY COLLECTION: 120,000 TITLES (8700 LINEAR FEET)  
MICROFICHE COLLECTION 564,500 TITLES (1215 LINEAR FEET)  
DAILY AVERAGE ASSUMES 20 WORKDAYS PER MONTH

**1. Information Currently Being Received From The  
Defense Technical Information Center By The Knox  
Library, Research Reports Division**

The RRD of the Knox Library currently receives government research reports, on microfiche, based on its DTIC profile. The profile states the types of reports desired by the Naval Postgraduate School (NPS), and may be updated at any time. NPS is a full distribution user, receiving all reports distributed by DTIC (except medical research reports).

The research reports are placed into an automated microfiche storage and retrieval system (a lektriever) by the library staff. These reports are stored sequentially by their accession document (AD) number; a serial number assigned to the report by DTIC upon initial receipt. The storage system contains approximately 500,000 microfiche reports. It is electronically operated by a librarian or a library assistant, who enters the location of the desired report into a control panel. The system electromechanically rotates the drawers of microfiche until the row containing the target fiche is accessible. Using an AD number the staff member physically searches the row.

**2. Problems With The Current Information System**

In the current information system the end-user of the information is isolated from its source. The access to the information database is limited to a key-word search which the student or faculty member prepares with the help

of a librarian. The librarian then logs onto an on-line DTIC database and searches the database using these key-word descriptions.

The product of this search is a printed listing of titles and authors of technical reports that has been retrieved using the key-words provided by the student or faculty member. The end-user analyzes the listing and selects reports that may be pertinent to his research and requests those reports from the librarian. The librarian then retrieves the selected reports from the RRD's DTIC microfiche holdings.

This multiple step procedure is time consuming, taking from several hours to several days to successfully complete. The steps of this retrieval process must be executed sequentially, with each step requiring staff intervention. Often the student or faculty member must return to the RRD on several occasions to complete a search. Given the system's design, it is highly impractical for the end-user to interact with the information directly.

The medium of microfiche, used by of the current information system, is also an impediment to the efficient utilization of the information database. Microfiche, while a relatively efficient storage method, is difficult to use. Flexibility available in retrieving information stored in this medium is limited. Retrieval of all records containing

**C. DESCRIPTION OF REQUIREMENTS THAT ARE NOT BEING MET BY THE CURRENT SYSTEM**

The Naval Postgraduate School is an institution providing technical and scientific education to commissioned officers of the Defense Department. The faculty are highly trained professionals, performing important research that has the potential to significantly change strategic, tactical, and operational aspects of the Navy and the Department of Defense.

Both the faculty and the officers attending the Naval Postgraduate School are potential users of the Knox Library's Research Reports Division, and of the DTIC database. Their time is an important resource that must be used effectively. Because of the importance of their mission, and their positions of responsibility, it is important and cost effective to provide these professionals with tools that optimize the use of their time.

Tools that provide optimal information access and handling capabilities are required to allow the most efficient utilization of time available for performing research. Researchers need tools that allow them to use their special knowledge in a given field to evaluate the applicability of research reports. Most of all, they need devices that allow rapid and timely access to information.

## **1. Additional Functionality Required In The DTIC Database**

The addition of two important functions would significantly increase the accessibility and value of the DTIC database. The first function is the capability to conduct full text searches of the database. This capability overcomes the limitations of indexing, which is directly related to the skill of the individual who created the indexes. For example, if a report was indexed under the term optical, then searching using the keys CD-ROM or CD-WORM would not find the report, unless these terms were explicitly included in the index. Full text searches enable researchers to broaden their query's scope to include the entire report's text. If the term, or combination of terms, specified are included in the text of any report in the information database then it will be identified as a potential source for the researcher. Second, the full text of selected reports should be available in a format that maximizes potential uses of the information, including printing, viewing on a terminal, electronic distribution, storage on a floppy disk, or editing the actual report. These capabilities should be available in an inexpensive form, preferably American Standard Code for Information Interchange (ASCII).

## **2. Present And Projected Workload And Capabilities Required In The DTIC Database**

The DTIC database can be expected to grow indefinitely. When planning the previously described enhancements to the system, its growth must be taken into consideration. Any system considered must have some reasonably easy method to augment the information database. The average growth experienced by this information resource is approximately 1,961 reports per month. Table 7 details the RRDs monthly transactions. Monthly or even quarterly updates to the database are acceptable. However, the media selected for storage must have the capability for unlimited growth.

Telecommunication facilities could further enhance the accessibility of the information database. The information's value would greatly increase if professional researchers could access it from their offices or even from their homes. The more accessible that the DTIC information database is to those performing DoD research, the greater its value will be to the research mission of the school. The lower the cost (in terms of time) of accessing information, then the greater the attractiveness of the option. However, data communications are beyond the scope of this paper and are deferred to future research.

Another important component is the dialogue management software. The systems software must be easy to use with minimal requirements for user training. The system



should have the capability for complex searches using boolean operators. This capability will maximize the researchers ability to find the information required. The ability to browse through selected reports should be available. Additionally, there should be a way to mark selected reports for copying onto a floppy disk, or even to download a selected report to another system. Optimally, the dialogue management system will allow the user to select a domain or sub-category of reports within which to perform more refined searches.

Ideally, the capability for multiple user access to the database will be available. Again, the easier the access to the information - the more it will be used and hence, the greater its value. Highly skilled professional researchers, in an optimal environment, should not have to queue up to access information.

#### **D. COMPATIBILITY LIMITED REQUIREMENTS**

##### **1. Federal Information Processing Standards**

All software, equipment, and material considered to meet the requirements stated in this document must be in accordance with specifications outlined in the Federal Information Processing Standards Publications (FIPS PUBS). There are three important reasons for this requirement. First, the importance of the information database as an investment and as a resource requires that it be afforded

the protection provided by adoption of recognized information processing standards. Second, the requirement for unlimited growth of an information database means that the media, equipment, and services required to support the database be available indefinitely. Finally, the requirements indicated under the provisions of Federal Government guidance require that FIPS PUBS standards be followed when selecting information processing material. The guidance cited in the National Technical Information Service's publication (1985) prescribing this action is quoted below.

Federal Information Processing Standards Publications (FIPS PUBS) are developed by the Institute for Computer Sciences and Technology (ICST) and issued under the provisions of the Federal Property and Administrative Services Act of 1949, as amended; Public Law 89-306 (79 Stat. 1127); Executive Order 11717 (38 FR 12315); and Part 6 of Title 15 of the Code of Federal Regulations (CFR).

## **2. Costs of Failure of Conversion**

The costs of any failure of conversion are basically the costs associated with the procurement of equipment and services for the transition to the new technology. This is potentially a very expensive conversion effort. Therefore, prototyping is recommended to allow the school to "buy" experience with conversion. In addition, it is further recommended that every attempt be made to collect a comprehensive set of reports on the successes and failures

of other Navy and government conversion efforts. This will allow sharing of lessons learned with other government activities, thereby reducing risk of conversion.

### **3. Steps to Be Taken to Foster Competition in Conversion**

The most important step that can be taken to ensure that competition is fostered to the maximum extent possible is to describe requirements in terms of established standards. Standards, as stated above, are available in FIPS PUBS. Standards are also available through Military Standards (MIL-STDS), the International Standards Organization (ISO), and the American National Standards Institute (ANSI). Description of requirements using established standards allows the greatest level of competition. These standards are available to the public and all vendors have the opportunity to produce products meeting the published standards. The use of established standards reduces the work required to specify government requirements.

### **4. Information Resources Contractors as Potential Sources for Satisfying Requirements**

A pre-solicitation survey was conducted to determine the availability of sources for meeting the requirements of this project. This was accomplished by publishing a Request For Information (RFI) in the Commerce Business Daily (CBD), a publication sponsored by the Department of Commerce to advertise Federal Government requirements. The announcement

appeared in the December 6, 1990 CBD, edition. The text of the publication is presented below.

Supply Officer, Naval Postgraduate School, Monterey, CA 93943 67 -- MICROFICHE READER-SCANNER/DIGITIZER Contact Barry Frew, 408/646-2392/Contracting Officer Hazel Rogers 408-646-2049. A microfiche reader-scanner, capable of accepting input from standard 24X, 98-image microfiche and digitizing the input with resolution of at least 151 pixels per mm and 151 scan lines per mm of actual fiche image. Signal to noise ratio of at least 20:1 is desirable. Automatic feed of microfiche is desirable. The microfiche reader-scanner should be capable of digitizing and transmitting data recorded on microfiche to, and interacting with, an IBM compatible PC/XT/AT microcomputer for storage of images on optical disc.

Numerous vendors replied to the RFI indicating that sufficient capabilities exist within the industry to create a contract for the full conversion project, or any subset thereof. A listing of the vendors that replied to the advertisement is presented in Appendix C.

Conversion projects that are ongoing in the government further support the existence of the capability within the industry for meeting these requirements. Similar projects currently underway include: the Navy Engineering Data Management Information and Control System (EDMICS) project (Kaebnick, 1990), and the Army Personnel Electronic Records Management System (PERMS) project (Lingvai, 1990). While this is not a comprehensive list of all current government microform-conversion projects, these two examples are fairly representative of the current activity in this field.

## **5. Parallel Operations of The Existing and the Conversion System**

Parallel operations of the current microfiche based system and the conversion system is essential until the new system has been proven. Validation is important for a number of reasons. First, the DTIC database represents an important source of research information that must be protected from any loss due to error or failure of any kind. Second, good information resource management practices indicate that a cross-over to new information processing services be effected only after the new services have been validated. Ideally, detailed testing and acceptance procedures should be specified. These procedures may be identified by reviewing test and acceptance reports from similar preceding projects.

## **E. RECORDS MANAGEMENT REGULATIONS**

The National Archives and Records Administration (NARA) has been designated as the executive agent for administration over the Federal records management program (NARA, 1990). They are the experts in the field of archiving information and in ensuring that the information will continue to be available to the government.

New electronic records created by the conversion process are to be managed in accordance with guidance published by NARA. There are several purposes addressed by these regulations. The first purpose is to ensure continued

availability of information required by components of the Federal Government. The second purpose is to ensure that information no longer required, or used, is disposed of properly. Finally, the security of sensitive information is also a concern under the regulation of NARA. All of these concerns must be addressed in the design of new systems.

#### **F. TRAINING REQUIREMENTS**

The essential training requirements must not be overlooked, or traded-off, for additional functionality or reduced costs. If compromises must be made, it is most strongly recommended that training be given the highest priority. The importance of training cannot be overstated. In order to maximize the value of the information database, an extensive and ongoing program of training for all categories of personnel must be included in the requirements. The value of the conversion of the DTIC database is the increased access to the valuable information it contains. This cannot be effected unless users understand how to use the new tools provided for them.

The following factors must be evaluated in determining the extent of the total investment that should be made in the training package.

1. The number of faculty, staff, and students that will be using the resource.

2. The level of education and skill, of these knowledge workers.

3. The value of these professionals' time.

Investments in training will be recovered through the better use of the Naval Postgraduate Schools' most valuable resource, the time of the professional staff, and the time of the professional officer-students. This is an application of the opportunity cost doctrine. That is, "The cost of inputs...are their values in their most valuable alternative uses" (Mansfield, 1982). When the costs of inputs associated with time consuming "hacking" and other trial and error approaches to training are considered, application of the opportunity cost doctrine illustrates that professional activities are the most valuable alternative use of a professionals' time.

#### **G. SPACE AND ENVIRONMENT**

Space requirements are basically of two types, space for conversion of the DTIC database, and space for operation of the new system. If the conversion is conducted off-site then space is not a consideration for conversion. If conversion is conducted on-site then ordinary office space should be sufficient. The office space should be located within the RRD. The space necessary for operations should be no greater than a standard office space. Associated facilities' costs must also be considered, including the

cost of utilities, building maintenance, and supplies. All of these cost elements should be addressed in the draft specifications distributed for comment to prospective vendors.

#### **H. CAPABILITY AND PERFORMANCE VALIDATION**

Two aspects of capability and performance must be considered. The first is the capability and performance of the conversion system. The second is the capability and performance of the final system delivered for use by the Naval Postgraduate School.

##### **1. Capability And Performance Of The System Used To Convert The DTIC Database From Microfiche To Optical Disc**

A primary concern in the conversion phase of the project is the quality of the raster image produced from scanning a microfiche image. Another concern is the time that it takes to convert an image to a digital format. Additionally, the quality of the raster image must be high enough to enable intelligent character recognition (ICR). Generally, the higher the density of pels per square inch (PPI) in the raster image the better and faster the ICR will be. However, the increased PPI is more expensive. Conversion time is also affected by the PPI used in scanning.

Regardless of issues involved with how character recognition is achieved, it is the most important criteria.



Character recognition will enable the production of American Standard Code for Information Interchange (ASCII) code from raster images. Therefore, the final capability and performance of the system selected must be in terms of successful character recognition. It is recommended that a specification requiring a 99.975 percent accuracy in the conversion of microfiche to ASCII characters be written into the draft set of specifications for comments from the vendor community.

**2. Capability And Performance Validation Of The Final System Delivered For Use In Retrieving Full Text Reports From The DTIC Database**

Capability and performance criteria should be consistent with applicable FIPS PUB, MIL-STD, ISO, and ANSI standards for optical systems. An ideal system should be capable of hosting multiple users. Because the number of users that the system should be capable of hosting simultaneously is a function of demand, and demand is unknown, a prototype system is advised to enable the school to "buy" that information. It is also recommended that the proper sizing of the final system be addressed by an expert in the field of Operations Analysis, perhaps as a thesis topic.

It is recommended that the initial prototype system be a single-user microcomputer. This technology is relatively inexpensive and is familiar to the majority of knowledge workers. It is further recommended that frequency

statistics be collected during the prototype phase to gain some profile of the demand for the new service. This single user system should be sized, in terms of CPU speed and memory, to enable maximum speeds available from the optical disc technology. As stated above, the access speeds of the optical disc should be in accordance with published standards.

## **I. SUMMARY OF REQUIREMENTS**

Requirements for the conversion of the current DTIC microfiche database located in the RRD of the Knox Library are listed below.

1. The microfiche records should be converted to an ASCII format, at a 99.975 percent level of accuracy.
2. The reports should be stored on an optical disc either ISO 9660 format (CD-ROM), ISO 9171 format (CD-WORM, 130mm), or in a CD 10 885 format (CD-WORM, 356mm). These are specified as candidate formats because they are the only optical disc standards that are currently in effect.
3. The access time to the reports on the disc should be the maximum speed specified as available in these standard formats. The systems response time should also be in accordance with FIPSPUB57, Guidelines For The Measurement Of Interactive Computer Service Response Time And Turn Around Time.
4. The dialogue management system should be in accordance with the specifications of section B., above.
5. The system should be capable of producing full-text retrieval of the research reports that can be distributed on floppy disk.
6. Graphics images should be available in a standard format such as the Computer Graphics Metafile (CGM), MIL-D-28003; and the Initial Graphics Exchange Specifications (IGES), MIL-D-28000.

7. Text files should be available in Standard Generalized Markup Language (SGML), MIL-M-28001.

**VIII. ANALYSIS OF ALTERNATIVES FOR MIGRATING UNCLASSIFIED  
TECHNICAL REPORTS FROM MICROFICHE TO OPTICAL DISC IN THE  
KNOX LIBRARY'S RESEARCH REPORTS DIVISION**

**A. THE NEED FOR AN ANALYSIS OF ALTERNATIVES**

Investments in information systems (IS) represent an important commitment of resources, both in time and money. Resources are expended for the procurement of information systems, for their maintenance, and for other related services in support of IS. Not to be forgotten, is the cost incurred through the use of IS, once it has been fully implemented. The initial investment is important but the enhancement of, or detracting from, productivity after the system is deployed is more significant.

Two concepts support the previous statements. First, considerable planning, capital investment, and implementation costs are expended to install a new information system, or to update an existing system. Second, once deployed, the new system will significantly impact the operations of an organization. This impact can be in three forms: 1) a significant increase in productivity (benefits received for value given); 2) no impact on productivity (no benefits received for value given); 3) or a decrease in productivity (benefits lost for value given).

All alternative information systems being considered must be thoroughly analyzed considering the above mentioned factors. The costs and benefits of each alternative must be reduced to a form that enables relative ease of comparison. The previously completed requirements analysis provides a basis for comparing and evaluating the costs and benefits of the proposed alternatives. (Haga and Lang, 1991)

#### **B. THE SIZE AND SCOPE OF THE ANALYSIS**

This analysis addresses forward-looking alternatives that have the potential capability of meeting the basic requirements of producing the DTIC technical reports (TR) in a digital, full-text format. The scope will be limited to technologies that have the capacity to store the volume of information in the DTIC database, and that currently have technical standards in place.

This analysis will address the conversion of the database from microfiche to a digital format and the installation of a system for retrieving and displaying the information in its digital form. It will not address issues beyond the Knox Library, Technical Reports Division's, DTIC database. It is narrowly scoped to optimize our focus on issues closely related to those concerning migrating information stored on microfiche to an optical storage environment.

## **C. INFORMATION OBTAINED CONCERNING THE MARKETPLACE**

### **1. Industry Contacts**

Numerous contacts within the microfiche scanning/digitizing marketplace were made by the authors when they attended two related trade shows. The first was the Multimedia Conference held in San Francisco, California on the 11th of October, 1990. The other was the Association of Records Managers and Administrators (ARMA) also held in San Francisco on the 5th of November, 1990. The most valuable aspects of attending these events was the opportunity to see information systems demonstrated and to ask questions of industry representatives.

Other industry contacts included site visits to the Terminal Data Corporation (TDC), where demonstrations of a full range of equipment were provided, as well as a tour of their manufacturing operations. Industry representatives from W. J. Schaffer, Co., Inc., and from Omni Micrographics visited the Naval Postgraduate School to inform the authors of their respective companies' abilities to meet the draft specifications published in the Commerce Business Daily (CBD) advertisement, listed above.

### **2. Contacts with Peer Groups**

Numerous government organizations are involved in moving their information databases into the optical storage environment. Contacts with these peer groups have been an important and useful aspect of our research. Valuable

information and experience have been readily shared by individuals in other organizations having similar interests.

The authors found four organizations that were of particular interest in the study. Each organization was involved in planning migrations of microfiche databases or document oriented information databases. These organizations were the Library of Congress (LOC), the Defense Technical Information Center (DTIC), the Navy Printing and Publication Service (NPPS), and the Army project management office for Personnel Electronic Records Management System (PERMS). The Library of Congress sponsored a pilot project for investigating the potential of migrating their collection to optical disc (Manns and Swora, 1987; Manns, 1990). DTIC has numerous initiatives in the field of optical storage that are ongoing, including a prototype containing over 20 years of technical report citations on CD-ROM. DTIC forwarded a copy of this to the authors for evaluation. The NPPS provided a copy to the authors of their requirements' analysis and analysis of alternatives for their directives issuance system. NPPS plans are to eventually place all Navy directives on optical disc. Finally the Army PERMS project office provided a copy of their Official Military Personnel Files Micrographics System Study to the authors. This document addressed the feasibility of migrating Army personnel records from microfiche to optical disc.

### **3. Published Materials**

Published materials were extensively utilized in obtaining information about the marketplace. Publications from numerous sources, including the government as well as the public press, were used in the familiarization process. The field of migrating microfiche to optical disc is just beginning to gather momentum, therefore, much of the information about this specific part of the optical technology was gained from in-house, and vendor publications. As discussed in chapters three and four, more generalized information about the fields of optical scanning and storage is widely available.

### **4. Sources of Information Available Through The Commerce Business Daily**

#### **a. Request for Information**

An advertisement placed in the Commerce Business Daily (CBD) by the authors proved to be a key source of information regarding migrating microfiche based information to optical disc. Numerous industry representatives replied to the advertisement (a listing of respondents is provided in Appendix C). The authors found these representatives to be very enthusiastic about their fields, and more than interested in providing information about the state-of-the-art in the field of migrating microfiche to optical disc.



**b. Solicitation of Comments on Draft Specifications**

A key recommendation for future research in this field, and for projects like the one evaluated in this thesis, is to solicit comments from industry on draft specifications prior to advertising a request for proposals (RFP). This will enable the creation of a "virtual brainstorming session" for fully defining the systems specifications. What we mean by a "virtual brainstorming session" is that by soliciting comments from industry, the government is in the position of being able to access some of the best minds in industry. A collection of ideas and comments from industry will help to produce a more comprehensive set of specifications. These ideas will enable a broad range of competition when the final solicitation for the project is advertised.

**D. IDENTIFICATION OF THE ALTERNATIVES**

The General Services Administration (GSA) has been tasked by Congress to implement the Brooks Act (Public Law 89-306). The Brooks Act outlines the basic policy for management of data processing equipment in the Federal Government. The Federal Information Resources Management Regulation (FIRMR) is the Federal Regulation issued by GSA in accordance with the Brooks Act.

The following alternatives must be evaluated in accordance with the FIRMR: non-information resources,

reconfiguring existing resources, mandatory programs and contracts, non-mandatory programs and assistance, sharing, in-house development, and contracting for new or additional services. Each of these alternatives will be discussed below. (General Services Administration, 1990)

#### **1. Non-Information Resource Alternatives**

The question explored by evaluating the alternatives of either maintaining the status-quo, or providing additional services that do not involve the use of IR must be addressed. If the status-quo is maintained, then no new or additional costs will be incurred. However, recurring costs associated with providing the service of the existing system must be considered. These include the cost of maintenance of equipment and services (such as dedicated data communications lines) required for providing these services. The cost of operating the RRD (including salaries, utilities, and facilities) is considered to be constant through out all alternatives. In accordance with the principles of economic analysis stated below, these expenses will not be considered in any of the alternatives evaluated in this analysis.

Any cost that will be incurred no matter what choice is made, any cost that must be borne regardless of the decision at hand, is not a cost of that particular choice or decision and need not be included in the analysis. (NAVDAC PUB 15, 1980)

The status quo, termed alternative one (ALT1) in this analysis, is not recommended because it will not meet the requirements previously identified. Faculty, staff, and students will not accrue any additional benefits from the existing system. Conversely, the argument presented in the requirements section stated that in fact there are hidden costs in lost productivity of the researchers, and a reduction in the value of the DTIC database because of the barriers to accessing the information. However, this alternative will be included in the analysis to illustrate the costs associated with the status quo.

Additional services could be provided in terms of newer microfiche readers, and facilities and staff for printing hard-copies of microfiche reports for the faculty and students. Again, as in the preceding paragraph, the authors argument is that these additional services will not remove sufficient barriers to the information to make this an attractive alternative. Therefore, this alternative will receive no further treatment in this analysis, and is dropped from consideration.

## **2. Reconfiguring Existing Resources**

The existing IR consists of a dedicated data communications line, terminals, and printers used to query the DTIC database. Reconfiguration of existing IR will not produce any significant increase in service. However, research into the options available for accessing the DTIC

database revealed that reconfiguring the existing system may produce cost savings over ALT1. This could be achieved by discontinuing the dedicated data communication services and implementing a dial-up data communications service. Because of the potential cost savings this option will be evaluated. This alternative will be termed alternative two (ALT2).

The only reconfiguration that could increase access to the information would be to allow end-users to dial into the DTIC Defense Research Development Test and Evaluation (DROLS) system themselves. It would be very difficult, if not impossible, to arrive at accurate estimates of the costs of providing this kind of service. This is primarily due to the "turnpike effect", i.e., it is difficult to predict usage of a service until it is made available. Because this alternative cannot be easily estimated, and it will not provide a significant level of increased access to the information base, it is eliminated from further consideration.

Another option that is available for reconfiguring existing resources is a fundamental change in the way citation information is obtained. So far analysis has focused on alternatives using some form of telecommunications to access the citation database on-line. However, DTIC recently announced a second prototype Compact Disc-Read Only Memory (CD-ROM) that contains 20 years of unclassified technical report (TR) citations (Defense Technical

Information Center, 1991). (DTIC's previous prototype contained six years of TR citations.) Because this is a stand-alone, microcomputer-based application, users do not have to be concerned with the problems associated with on-line systems, e.g., telecommunications problems, computer down time, and operational hours.

In the requirements analysis it was determined that the RRD required both classified and unclassified citations. The DTIC CD-ROM offers only unclassified citations, therefore, to employ this option the RRD would have to maintain some form of on-line capability.

An alternative considered by the authors is to use the CD-ROM to the greatest extent possible, and to access the on-line system via dial-up lines, on an "as-needed" basis. This alternative has the potential to significantly reduce telecommunication's costs. Consideration will be given to this option, and is termed alternative three (ALT3).

The alternatives considered thus far only partially address the requirements as stated earlier in chapter seven. Alternatives one through three address the status quo, and suggest slight improvements that would increase its cost effectiveness. They have not addressed the issue of producing the technical reports in a digital format with a full-text retrieval capability.

The authors will now focus on the requirement of producing the TRs in a digital format, with a full-text retrieval capability. This will be the central focus of the remaining alternatives. In the next alternative, termed alternative four (ALT4), a change in policy is introduced as a low-cost method of eventually achieving a digital format, with full-text retrieval, in the RRDs holdings of DTIC Trs. The proposed policy changes the acquisition of microfiche Trs to the acquisition of all new digital-format Trs. Employing this alternative will gradually move the RRD toward a full-text TR information base.

Alternative four, and all remaining alternatives, will also include the basic components of ALT3. That is, they will all employ the DTIC TR citations on CD-ROM and dial-up, on-line TR citation service on an "as-needed" basis.

### **3. Mandatory and Non-Mandatory Programs and Contracts**

General Services Administration (GSA) mandatory-for-use programs must be evaluated in considering alternatives for meeting requirements for new information systems. These programs include a number of government-wide programs that are required. One required program that must be considered is the excess IR equipment program. This is a program that promotes the reuse of government equipment that is no longer called for. This potential source of equipment may be checked by contacting GSA's Authorization Branch.

Other sources that must be evaluated are GSA mandatory-for-use contracts, non-mandatory contracts, as well as other existing government contracts that may be applicable. These programs are not applicable in the present analysis, because initiatives in this field within the government are just beginning and the kinds of equipment required are not yet available via mandatory or non-mandatory sources (Black, 1991). Therefore, they are eliminated from further consideration. (General Services Administration, 1990)

#### **4. Sharing Excess Capabilities of Other Federal Agencies**

Sharing involves identifying other federal agencies that have similar on-going projects, and that have the scope of sharing excess capabilities in their contracts (Black, 1991). The purpose of this alternative is to encourage agencies to share additional capabilities that are not fully utilized, or to combine requirements to reduce the total overall cost to the government. GSA provides assistance in identifying opportunities for sharing IR resources.

This is considered to be a viable alternative for effecting the migration of the RRD's DTIC holdings to an optical environment. In the course of our interviews with the Director of the Knox Library's Research Reports Division the authors learned of several DTIC initiatives in the field of optical storage. The possibility of resource sharing

with DTIC, or perhaps even acting as a beta test site are very attractive alternatives. However, those kinds of initiatives are only in their early planning stages at DTIC and are not mature enough to be considered in this analysis. (Jones, 1990)

The authors have identified several Navy contracts for migrating information databases from microfiche to optical storage environments. Utilization of one of these contracts will be considered as a viable alternative. Within this option, two alternatives will be considered. The first is a partial conversion of the RRD's DTIC holdings, including the most recent five years of the information base; this option is termed alternative five (ALT5). The second is a full conversion of the RRD's DTIC holdings; this option is termed alternative six (ALT6).

#### **5. In-House Development**

Criteria that should be considered in evaluating in-house development are the number of technically qualified personnel that are available. This is a high risk alternative especially if there is no previous experience in the technical area being addressed.

This is true in the case of migrating the RRD's DTIC microfiche database to optical storage. There are no personnel available for this project with the specific technical expertise needed. In the project being considered in this paper, specific technical expertise is required in



the areas of microform scanning, intelligent character recognition, and in indexing the information base. Errors in these areas could render the information base useless and result in a loss of the investment. Therefore, this alternative will not be considered in our analysis.

#### **6. Contracting for New or Additional Services**

New or additional services contracting is the last alternative that should be considered, for several reasons. First, this is the most time consuming alternative. It requires development of detailed specifications, synopsis in the Commerce Business Daily, evaluation of vendor proposals, and potential arbitration of contract action protests. Secondly, this is an expensive alternative because of the administration required to establish a new contract and to manage it properly. Finally, this alternative contains the greatest risk to the government. The risk is one associated with the establishment of a new contract for equipment and services that does not have a demonstrated success record. (General Services Administration, 1990)

This alternative will not be considered in this analysis because there are other viable alternatives to be considered. Namely, the alternatives of utilizing a previously established contract, and the alternative of sharing the expense of this information system development with DTIC.

## **7. A Summary of the Alternatives Proposed**

In the preceding process of identifying options to be considered in this analysis six choices were proposed. They are listed below for easy reference.

1. Alternative one (ALT1), status quo
2. Alternative two (ALT2), reconfiguring data communications lines to yield a more cost effective operation
3. Alternative three (ALT3), using the DTIC TRs on CD-ROM with dial-up data communication lines on an "as-needed" basis
4. Alternative four (ALT4), using the DTIC TR CD-ROM with dial-up data communication lines on an "as-needed" basis, with a policy change to begin electronic document acquisition
5. Alternative five (ALT5), using the DTIC TR CD-ROM with dial-up data communications, and a partial conversion of the RRD's DTIC holdings (the most recent five years of data)
6. Alternative six (ALT6), using the DTIC TR CD-ROM, with dial-up data communications, and a full conversion of the RRD's DTIC holdings.

Two objectives were considered in developing these alternatives. The first purpose was to attempt to develop a comprehensive list of alternatives to address the requirements identified earlier. The second goal was to structure the alternatives in such a way as to offer a range of choices. By offering a range of choices, "all-or-none" decisions can be avoided. Thereby a continuum of choices, in terms of degree of change and costs, are provided to the decision-maker.

During the process of structuring the alternatives, the authors determined that the range of choices generated could be divided into two decisions. The first decision is to choose between alternatives one through three, and the second decision is to select from alternatives four through six.

Decision one and decision two are distinguished from one another by the comprehensiveness of the solution prescribed. Decision one addresses only a partial solution, i.e., it addresses improving the methods of searching for citations. Decision two addresses the requirement for converting the RRD's DTIC holdings to a digital format. Decision one does not require the selection of any of the choices in decision two. The decision-maker can elect to adopt one of the choices in decision one and decide not to convert the RRD's DTIC holdings to a digital format. However, decision two assumes the selection of alternative three, and offers a range of alternatives that allow conversion of the RRD's DTIC holdings to a digital format. Alternative three is assumed in decision two because during the conversion of the RRD's DTIC holdings from microfiche to a digital format a method of searching those citations that have not been converted to a digital format will be required.

## **E. DETERMINING THE MOST ADVANTAGEOUS ALTERNATIVE**

### **1. Cost Factors**

The FIRMR requires Federal agencies to prepare a cost analysis of each feasible alternative, using the present value of money, when the value of the acquisition is expected to be greater than \$50,000 (General Services Administration, 1990.) Haga and Lang (1991) explain that present value analysis is a method of placing the alternatives under examination on an equal basis, as of the date they are compared. The cost analysis should consider all sources of expense including both one time and recurring costs. Sources of expenditure that must be considered are conversion, personnel, supplies, energy, maintenance, space, administrative costs of contracting, and contract prices.

Conversion costs are those expenses related to conversion, replacement, or disposal of existing software. Conversion costs do not apply to the DTIC technical reports (TR) database as it is currently implemented in the RRD of the Knox Library, and as such will be dropped from further consideration.

Costs associated with the basic operation of the RRD, such as personnel and the cost of the facility, are constant costs throughout all of the alternatives considered, and therefore (as previously discussed) will be disregarded in the analysis. Each of the other factors listed above do pertain to the problem being analyzed.

Tables 8 and 9 exhibit costs associated with the relevant factors for each alternative being evaluated. Table 8 presents the alternatives associated with decision one and Table 9 presents alternatives associated with decision two.

## **2. Non-Cost Factors**

The purpose of evaluating non-cost factors is to ensure that the specifications outlined in the requirements section are adequately addressed, and to evaluate benefits to be gained by the government in adopting one of the systems being evaluated. A key concern in analyzing a given alternative is its "value to the government" in reducing cost and increasing capability.

There are two kinds of non-cost factors to be considered in an analysis of alternatives. They are functional factors and risk factors. The functional factors are the benefits that should be derived from a system. The requirements analysis outlines these benefits and should be addressed. Risk factors are elements that could possibly prevent the achievement of the objectives stated in the requirements analysis. They are analyzed to aid in determining the probability of the successful achievement of the objectives stated in the requirements analysis. A GSA publication, A Guide For Requirements Analysis and Analysis of Alternatives (1990) fully describes the specific functional and risk factors recommended for inclusion in an analysis of alternatives. This analysis will address only

TABLE 8. DECISION NUMBER ONE COSTS

CURRENT SYSTEM	ALT1: STATUS: QUO					ALT2: RECONFIGURE IR					ALT3: CD-ROM/DIAL-UP				
	YEAR1	YEAR2	YEAR3	YEAR4	YEAR5	YEAR1	YEAR2	YEAR3	YEAR4	YEAR5	YEAR1	YEAR2	YEAR3	YEAR4	YEAR5
TELECOMMUNICATIONS															
DEDICATED PHONE LINE	15600	15600	15600	15600	15600	0	0	0	0	0	0	0	0	0	0
LINE USAGE	0	0	0	0	0	1000	1000	1000	1000	1000	500	500	500	500	500
DTIC ACCESS	0	0	0	0	0	4478	4478	4478	4478	4478	2239	2239	2239	2239	2239
TERMINAL MAINTENANCE	2570	2570	2570	2570	2570	90	90	90	90	90	90	90	90	90	90
TOTAL	18170	18170	18170	18170	18170	5568	5568	5568	5568	5568	2829	2829	2829	2829	2829
TOTAL COST	18170	18170	18170	18170	18170	5568	5568	5568	5568	5568	2829	2829	2829	2829	2829
NEW SYSTEM COSTS															
ADMINISTRATION COSTS															
CONTRACT COSTS															
PCs (2000 EA)															
OPTICAL DRIVE(600 EA)															
CONVERT MICROFICHE															
DISCS (WORM)															
JUKEBOX															
SOFTWARE(RETRIEVAL)															
SOFTWARE(DRIVER)															
SOFTWARE(NETWORK)															
MAINTENANCE/SUPPLIES															
TOTAL	0	0	0	0	0	0	0	0	0	0	1000	3600	1000	1000	1000
TOTAL COST	18170	18170	18170	18170	18170	5568	5568	5568	5568	5568	3829	6429	3829	3829	3829

## ASSUMPTIONS:

- Alt. 1 requires \$1300 per month for a dedicated phone line to DTIC  
 Alt. 1 requires annual maintenance for dedicated terminal  
 Alt. 2 requires 4 hours of long distance dial-up service per day at \$14/hr  
 Alt. 2 requires 149 hours of connect time at \$30/hr  
 Alt. 2 requires annual maintenance on 1 STU-III secure phone  
 Alt. 3 requires an annual subscription fee from DTIC for the CD-ROM  
 Alt. 3 requires annual maintenance and supplies for the CD-ROM system.  
 Alt. 3 requires annual maintenance on 1 STU-III secure phone  
 Alt. 3 decreases dial-up usage by 1/2 due to increased usage of the CD-ROM

TABLE 9. DECISION NUMBER TWO COSTS

NEW SYSTEM	ALT4: CD-ROM/DIAL-UP/POLICY CHANGE					ALT5: CD-ROM/DIAL-UP/PARTIAL CONVERSION					ALT6: CD-ROM/DIAL-UP/FULL CONVERSION				
	YEAR1	YEAR2	YEAR3	YEAR4	YEARS	YEAR1	YEAR2	YEAR3	YEAR4	YEARS	YEAR1	YEAR2	YEAR3	YEAR4	YEARS
ADMINISTRATOR	0	0	0	0	0	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
PCS/WORKSTATIONS	0	8000	4000	0	0	8000	0	0	0	0	20000	0	0	0	0
OPTICAL DRIVES	0	1200	1200	0	0	1200	0	0	0	0	1200	0	0	0	0
CONVERT MICROFICHE	0	0	0	0	0	23000000	0	0	0	0	23000000	23000000	23000000	23000000	23000000
JUKEBOX	68000	0	0	0	0	110000	0	0	0	0	110000	0	0	0	0
SOFTWARE(RETRIEVAL)	15400	1400	1400	1400	1400	15400	15400	15400	15400	15400	15400	15400	15400	15400	15400
SOFTWARE(DRIVER)	5000	500	500	500	500	30000	500	500	500	500	30000	5000	5000	5000	5000
SOFTWARE(NETWORK)	20000	500	500	500	500	20000	500	500	500	500	20000	2000	2000	2000	2000
MAINTENANCE/SUPPLIES	5000	5000	5000	5000	5000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
TOTAL	113400	16600	12600	7400	7400	23236600	56400	56400	56400	56400	23236600	23062400	23062400	23062400	23062400

## ASSUMPTIONS:

- Alts. 4, 5 and 6 require two CD-ROM drives  
 Alts. 4, 5 and 6 require four microcomputers or workstations  
 Alts. 4 requires a WORM drive(s) capable of storing 50 gb  
 Alt. 5 requires conversion of 5 years of RRD microfiche  
 Alts. 5 and 6 require a full time project manager  
 Alt. 6 requires conversion of 25 years of RRD microfiche  
 Alts. 5 and 6 require a WORM drive(s) capable of storing 100 and 250 gb respectively

the functional factors. Risk factors are entrusted to a future study.

**F. THE DECISION PROCESS: SELECTING AND REPORTING THE MOST BENEFICIAL ALTERNATIVE**

The end-product of the analysis of alternatives is a substantive demonstration of the decision process. This is usually in the form of a tabular presentation of the results of the decision techniques used to support the final recommendations. Several methods are specifically recommended by the General Services Administration (GSA) for economic analyses. These include present value (PV) analysis and benefit-cost ratio (BCR) analysis. Haga and Lang (1991) have issued a publication entitled Economic Analysis Procedures for ADP, that outlines how to apply the procedures identified by GSA, utilizing a step-wise methodology.

These techniques of economic analysis will be described and applied to the decisions under study in this paper. Explicitly stated, the objective of this exercise is to determine which of the alternatives addressed in this report, are the most advantageous for the Naval Postgraduate School's, Knox Library, Research Reports Division.

**1. The Present Value (PV) Analysis**

Present value analysis is a technique used to express each alternative in equal terms. It allows the analyst to place alternatives on a level field in terms of



time and cost. (Haga and Lang, 1991) The reasons that present value analysis is necessary are best defined in the GSA publication A Guide for Requirements Analysis and Analysis of Alternatives, as cited below.

Benefits accruing in the future are worth less than the same level of benefits that accrue now; and Costs that occur in the future are less burdensome than costs that occur now. (GSA, 1990)

Present values are computed by applying a discount factor to the costs, and to the benefits when they are quantifiable. This procedure, termed discounting, consists of multiplying the factors being considered by a discount factor. Discount factors are published by the Office of Business and Management in OMB Circular No. A-94. Tables 10 and 11 display the present value analysis for this project. Table 10 addresses the alternatives for decision one, and Table 11 addresses the alternatives for decision two.

## **2. The Benefit-Cost Ratio (BCR) Analysis**

An important concern in evaluating alternative investments is whether or not they will yield benefits commensurate with the costs. The BCR is a tool to measure the relative value of alternatives. This tool is an indicator of the benefits gained for each dollar spent. The alternative with the highest BCR is the most cost effective. There are two different situations in which BCR may be applied. One is when benefits are quantifiable and the

TABLE 10. PRESENT VALUE ANALYSIS, DECISION NUMBER ONE

	YEAR1	YEAR2	YEAR3	YEAR4	YEAR5
ALT1: STATUS QUO					
ANNUAL COSTS	18,170	18,170	18,170	18,170	18,170
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	17,334	15,753	14,318	13,028	11,847
5 YEAR TOTAL: \$72,280					
ALT2: RECONFIGURE IR					
ANNUAL COSTS	5,568	5,568	5,568	5,568	5,568
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	5,312	4,827	4,388	3,992	3,630
5 YEAR TOTAL: \$22,150					
ALT3: CD-ROM/DIAL-UP					
ANNUAL COSTS	3,829	6,429	3,829	3,829	3,829
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	3,653	5,574	3,017	2,745	2,497
5 YEAR TOTAL: \$17,486					

TABLE 11. PRESENT VALUE ANALYSIS, DECISION NUMBER TWO

	YEAR1	YEAR2	YEAR3	YEAR4	YEAR5
ALT4: CD-ROM/DIAL-UP AND POLICY CHANGE					
RECURRING	113,400	16,600	12,600	7,400	7,400
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	108,184	14,392	9,929	5,306	4,825
5 YEAR TOTAL: \$142,635					
ALT5: CD-ROM/DIAL-UP AND PARTIAL CONVERSION					
ANNUAL COSTS	23,236,600	56,400	56,400	56,400	56,400
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	22,167,716	48,899	44,443	40,439	36,773
5 YEAR TOTAL: \$22,338,270					
ALT6: CD-ROM/DIAL-UP AND FULL CONVERSION					
RECURRING	23,236,200	23,062,400	23,062,400	23,062,400	23,062,400
DISCOUNT FACTOR	0.954	0.867	0.788	0.717	0.652
DISCOUNTED COSTS	22,167,716	19,995,101	18,173,171	16,535,741	15,036,685
5 YEAR TOTAL: \$91,908,414					

other is when benefits are not quantifiable. Each of these situations are discussed below.

**a. The BCR When Benefits Are Quantifiable**

If projects have the objectives stated in terms of required outputs, then benefits are relatively easy to quantify. In these cases the appropriate formula to use is:  $BCR = \text{Quantifiable Output Measure} / \text{Uniform Annual Cost}$ . Examples of quantifiable output measures include miles per gallon, dollars per horse-power, or dollars per megahertz. The uniform annual cost (UAC) method accounts for both the time value of money, and for the differing time spans in the economic lives of the options evaluated. It places all alternatives on a level field to enable valid comparisons of alternatives. (Haga and Lang, 1991)

This technique will not be used in this analysis because the benefits are non-quantifiable. The potential value to be received from the alternatives in this analysis are increased functionality and capability. These may result in greater service to the RRD's patrons.

**b. The BCR When Benefits Are Not Quantifiable**

The greatest difficulty in applying the BCR technique is in quantification of the benefits. The BCR technique is a very versatile methodology in that it can still be applied when precise quantification of the benefits is not possible. Due to the fact that this method requires

a degree of subjectivity, the analyst must include the rationale used in determining the aggregate benefit values.

Aggregate benefit values are usually derived by employing techniques using weighted or scaled values (similar to a Likert scale) to derive the benefit values (Haga and Lang, 1991). The formula for the BCR when the benefits are non-quantifiable is:  $BCR = \text{Aggregate Benefit Value} / \text{Uniform Annual Cost}$ . This technique will be used because precise quantification of the benefits is not possible.

The methodology used to derive the benefit factors and their weighted values was a three step procedure. First, the authors "brainstormed" all of the benefits factors within each alternative. Second, the survey depicted in Appendix F was developed by the authors, with the aid of the director of the Knox Library and one of his key staff members. The survey was given to the directors of the library and to all staff members who utilize library information systems when performing their duties.

Table 12 represents the benefit weights and rankings for each alternative under consideration. The functional factor weights (WT), located in the first column in the table, depict the results of the survey (represented as an average weight.) The aggregate benefit value (ABV)

derived for each alternative evaluated can then be used to calculate the BCR using the method described above.

#### **G. A DISCUSSION OF THE RESULTS OF THE ECONOMIC ANALYSIS**

As previously mentioned, the alternatives were divided into two decisions, decision one and decision two. Figure 3 graphically illustrates the two levels of decisions that can be made based on this economic analysis. Decision one contains the status quo and two additional alternatives that use graduated levels of new technology, to access citation information. Decision two contains alternatives using three different levels of the same advanced technology to produce the technical reports in a digital format with the capability for full-text retrieval.

##### **1. The Evaluation of Decision One**

Decision one is focused on alternatives for obtaining citations from the DTIC technical reports (TR) database. Data communications costs and the costs associated with implementing a CD-ROM system are the key elements to be considered when exploring ways to improve access to technical report citations. Table 13 summarizes the relevant decision aids that are available to assist the decision-maker, in decision one. It displays the resultant aggregate benefit value (ABV) analysis, the present value (PV) analysis, and the benefit cost ratio (BCR) analysis.

TABLE 12. BENEFIT WEIGHTS AND RANKINGS

FUNCTIONAL FACTORS	WT	ALT1 ADJ		ALT2 ADJ		ALT3 ADJ		ALT4 ADJ		ALT5 ADJ		ALT6 ADJ	
ACCEPTANCE	8	2	16	2	16	6	48	9	72	9	72	9	72
ACCESSIBILITY	8	1	8	1	8	6	48	8	64	10	80	10	80
ACCOUNTABILITY	8	2	16	2	16	3	24	9	72	9	72	9	72
AVAILABILITY	8	2	16	2	16	5	40	9	72	9	72	9	72
CONNECTIVITY	4	1	4	1	4	1	4	9	36	9	36	9	36
EXPANDABILITY	4	4	16	4	16	4	16	8	32	8	32	8	32
FLEXIBILITY	5	3	15	3	15	5	25	9	45	9	45	9	45
MAINTAINABILITY	7	4	28	7	28	8	56	8	56	8	56	8	56
MATURE TECH.	8	9	72	9	72	9	72	7	56	7	56	8	56
PRODUCTIVITY	9	3	27	3	27	4	36	9	36	9	36	10	90
QUALITY OF SEARCH	9	5	45	5	45	8	72	9	81	9	81	9	81
RELIABILITY	8	3	24	4	32	6	48	8	64	8	64	8	64
SECURITY	3	3	9	3	9	3	9	9	27	9	27	9	27
STAFF MORALE	6	3	18	3	18	7	42	8	48	9	54	10	60
USER FRIENDLINESS	7	3	21	3	21	7	49	9	63	9	63	10	70
TOTAL		319		341		541		788		819		849	

Notes:

Columns headed with "ALT" contain functional factors scores  
Columns headed with "ADJ" contain the weight adjusted scores

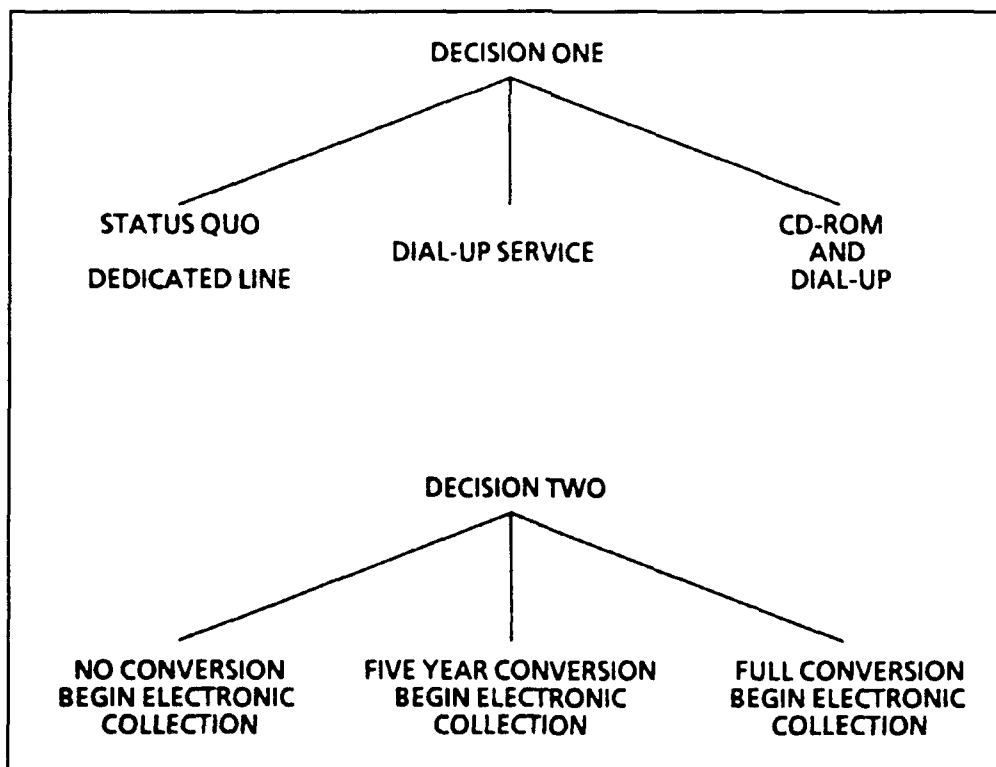


Figure 3. Decisions Available to the Knox Library Research Reports Division

The ABV of ALT3 is significantly greater than for the other two alternatives. The PV of ALT3 is also lower than the other two alternatives, and significantly lower than ALT1, the status quo. The BCR of ALT3, as expected, is significantly larger than either of the other alternatives in decision one. This analysis indicates that greater value and benefits can be achieved, at lower costs, by electing alternative three, of decision one.

## **2. The Evaluation of Decision Two**

The focus of decision two is on the rate and degree to which microform technical reports are converted to a digital format. ALT4 proposes drawing a baseline at the current point in time, deciding to collect all future TRs in a digital format, and thereby gradually achieve the objective of having the most recent TR database in a digital format. ALT5 proposes converting the most recent five years of TRs now, and collecting all future TRs in a digital format. ALT6 proposes a full conversion of the complete RRD DTIC holdings to a digital format now, and collecting all future reports in a digital format.

The benefits attributable to having information in a digital format are significant and so are the costs. The three alternatives provide varying degrees of conversion of existing microfiche, while all three have the intent of achieving a full-text, digital format for current technical reports.

Table 14 provides a summary of the pertinent decision aids available to assist the decision-maker in decision two. It displays the aggregate benefit value (ABV) analysis, the present value (PV) analysis, and the benefit cost ratio (BCR) of each of the three alternatives in decision two. There is little difference between the ABV of the three alternatives, but the PV variance is significant. The PV of ALT4 is significantly lower than the other two alternatives in decision two. Because the ABV for the three alternatives is relatively equal and the variance between the PVs is great, it is expected that the alternative with the lowest PV costs will have the greatest BCR value. In fact, the BCR analysis determined that ALT4 may yield the greatest value for the investment.

### **3. The Value of Information**

One factor which must weigh heavily in any decision to convert technical reports stored on microfiche is the underlying value of the information. While research reports certainly do have a high initial value, this value decreases over time. Decision-makers must determine which information is valuable enough to convert and maintain online. Dated information, that may be accessed less frequently, may not warrant the expense of conversion to a digital format.

To determine the value of the information, decision-makers must turn to the end-user of the technical



TABLE 13. BENEFIT/COST RATIO ANALYSIS, DECISION ONE

BENEFIT/COST ANALYSIS	AGGREGATE BENEFITS	PV COSTS (000)	BENEFIT/ COST RATIO
ALT1: STATUS QUO	319	72	4
ALT2: RECONFIGURE IR	341	22	15
ALT3: CD-ROM/DIAL-UP/NO POLICY CHANGE	541	17	31

TABLE 14. BENEFIT/COST RATIO ANALYSIS, DECISION TWO

	AGGREGATE BENEFITS	PV COSTS (000)	BENEFIT/ COST RATIO
ALT4: CD-ROM/DIAL-UP/WITH POLICY CHANGE	788	143	6
ALT5: CD-ROM AND PARTIAL CONVERSION	819	22,338	0.037
ALT6: CD-ROM AND FULL CONVERSION	849	91,908	0.009

reports. It is recommended that additional information be collected from the consumers of the technical reports, via surveys, to determine the demand for the different types and ages of technical reports. Demand data can aid in structuring the TR database conversion decision regarding which reports to convert, and when to convert them, given the limited resources.

## **IX. CONCLUSION AND RECOMMENDATIONS**

### **A. CONVERSION TO FULL-TEXT FORMAT IS POSSIBLE**

Advances in optical technology have made it possible to maintain large information bases in a character coded format. Full-text search and retrieval software developments have made it possible to increase the accessibility and, therefore, the value of the information contained in these large information bases. The combination of these two technologies has increased the interest in converting existing microfiche files to optical storage media. The technology to convert existing microfiche files is well developed and there are many organizations that specialize in providing conversion services, however, the decision to undertake a backfile conversion is by no means a trivial one.

### **B. THE DISCIPLINE OF ECONOMIC ANALYSIS SHOULD BE USED**

The advantages of having full-text search capabilities must be weighed against the costs of conversion. While the costs of conversion are easily quantified, the benefits to be derived from such a conversion are less so. Factors such as value of researchers time, frequency of access to documents, and the value of specific documents can help in arriving at an objective cost benefit figure. However, such

intangible factors as obsolescence, connectivity, and increased functionality must also be considered. For many technologically oriented organizations the ability to thrive in a dynamic technological environment is a critical success factor and building an infrastructure for dealing with such change should be considered in the decision.

Each organization must follow an economic analysis discipline to examine the factors that influence the conversion decision in its specific case. The decision-maker must decide which course of action is best for the organization after the costs and benefits have been analyzed. An economic analysis does not make this decision for him, rather it provides an input to his decision-making process. The true value of the discipline of economic analysis is that it requires an explicit statement of the costs and benefits of various alternatives as well as underlying assumptions. The decision-maker can then evaluate the relative importance assigned to various factors as well as the reasonableness of the assumptions. By bringing these factors out into the open, the economic analysis enables better decision making.

#### **C. KNOX LIBRARY RESEARCH REPORTS DIVISION RECOMMENDATIONS**

An application of the discipline of economic analysis to the Knox Library RRD made it apparent that there were two distinct decisions involving the use of optical technology

to improve service. The first involved access to bibliographic citations while the second involved the bigger issue of access to the full-text of technical reports.

#### **1. Optical Technology to Improve Citation Access**

Of the three alternatives affecting access to the technical report citations, the CD-ROM option proved to be the dominant alternative. Conversion to a dial-up means of access to citation information in lieu of the existing dedicated line will yield more than enough savings to cover the costs of acquiring the CD-ROM system to complement the dial-up capability. In addition to added functionality provided by CD-ROM, the implementation of this system will serve as a first step toward developing optical storage expertise in the Knox Library.

#### **2. Optical Technology to Improve Full-Text Access**

Three alternatives related to improving access to the full-text of technical reports highlight the large expense of backfile conversion. The conversion process is simply not yet fully automated and is, therefore, expensive. However, the advantages of full-text search and retrieval remain attractive and are worth pursuing. For that reason, the alternative that calls for no backfile conversion, but ultimately achieves a full-text storage and retrieval system is recommended. By investing in small scale prototypes for electronic document acquisition, storage, and retrieval, the Naval Postgraduate School can make a valuable contribution

to applied research as well as position itself to take advantage of future full-text retrieval opportunities.

While large-scale backfile conversion is not a feasible alternative for a single site such as Naval Postgraduate School, it may prove to be feasible at a higher organizational level. The Defense Technical Information Center should continue to investigate the issue of converting to a full-text storage and retrieval system, perhaps involving Naval Postgraduate School as a beta test site. Existing DTIC projects in both CD-ROM and full-text retrieval indicate interest in improving access to DTIC's technical reports and future cooperation with NPS in this area is recommended. Economies of scale, lower distribution costs, and ability to acquire necessary expertise are all factors which suggest DTIC as the logical initiator for such conversion projects.

#### **D. CONCLUSION**

Full-text storage and retrieval systems provide a cost effective way of dealing with the growing problem of information overload. If an organization is to take full advantage of this technology, it must begin now to establish policies and infrastructures that will allow migration to optical-based, full-text retrieval systems without an expensive backfile conversion process. Developing electronic document acquisition standards and gaining

experience in the field of optical storage and retrieval systems must be given priority. Planning and budgeting for these programs now will certainly yield long-term cost savings and benefits. The future of document storage and retrieval lies in full-text retrieval systems and those organizations that prepare now will reap the biggest rewards.

## APPENDIX A

### CHECKLIST FOR ASSESSING SOFTWARE RETRIEVAL CAPABILITIES

#### I. User Interface

- A. What impression does the overall interface make?
- B. Is the interface designed for one or more user levels (novice/expert)? Is it menu-driven, command-driven or a combination?
- C. Are function keys used clearly and appropriately?

#### II. Screen Displays

- A. Are screen displays clear and well organized?
- B. Do they make effective use of color, graphics, windowing, special features?
- C. Is the display information appropriate for the intended audience?

#### III. Retrieval Modes

- A. What search features are offered?
  - 1. Boolean operators? Which ones? Is logic implicit, by command or a combination?
  - 2. Positional operators?
  - 3. Nested logic?
  - 4. Field qualification? How is it specified?
  - 5. Wild-card symbols and truncation: Number of characters specified or open?
- B. Can search strategies be modified easily?
- C. Are search statistics clearly displayed?
- D. Can search strategies be saved and re-executed?
- E. Does the system have an on-line thesaurus? Is it quickly and easily available? What are the protocols for entering controlled language terms?

#### IV. Response Time

- A. How does the response time compare to that of other media? With that of other optical systems?



B. Are appropriate processing messages displayed?

C. Is there a break function?

V. Post-Processing Capabilities

A. Displaying? Can formats be selected, altered?

B. Printing? Can citations be viewed first? Can formats be selected, changed? Do default formats include all important information?

C. Downloading? Can text be saved to disc or diskette? Can files be reformatted edited, sorted? Are results compatible with popular software programs?

D. Can default settings for format be changed? Can limits be placed on the number of citations that can be printed or downloaded?

VI. On-Screen Help

A. Are help screens readily available from any point in search?

B. Is the information presented on the help screens clear, concise, effective?

VII. Documentation

A. What documentation is supplied with the system? User manual, reference cards, templates, posters?

B. Are the materials clear, well-illustrated, up-to-date with system capabilities?

C. If more than one company is involved, what are the responsibilities of each?

D. Is toll-free telephone assistance provided? During what hours?

(Eaton, McDonald, and Salve, 1989)

## APPENDIX B

### REGULATIONS FOR INFORMATION RESOURCE MANAGEMENT

#### FEDERAL REGULATIONS

I. There are four regulations implementing the public laws

- A. Federal Acquisition Regulation (FAR)
- B. Federal Information Resources Management Regulation (FIRMR)
- C. DoD FAR Supplement (DFARS)
- D. Agency Supplement Regulations
  - 1. Navy Acquisition Procedures Supplement (NAPS)

II. DoD Directives and Instructions

- A. DoDD 4105.62, Selection of Contractual Sources for Major Defense Systems
- B. DoDD 4120.3, Defense Standardization and Specification Program
- C. DoDD 5000.1, Major and Non-Major Defense Acquisition Programs
- D. DoDD 5000.29, Management of Computer Resources in Major Defense Systems
- E. DoDI 5000.31, Interim List of DoD Approved High Order Programming Languages
- F. DoDD 5200.28, Security Requirements for Automated Information Systems
- G. DoDD 7740.1, DoD Information Resources Management Program
- H. DoDD 7740.2, Automated Information System Strategic Planning
- I. DoDD 7920.1, Life Cycle Management of Automated Information Systems
- J. DoDD 7930.1, Information Technology Users Group Program

- K. DoDI 7930.2, ADP Software Exchange and Release
- L. DoD 7950.1-M Defense Automation Resources Management Manual

### III. Navy Department Instructions

- A. SECNAVINST 5000.1C, Major and Non-Major Acquisition Programs
- B. SECNAVINST 5200.32, Management of Embedded Computer Resources in the Department of the Navy Systems
- C. SECNAVINST 5231.1, Lifecycle Management Policy and Approval Requirements for Information Systems Projects
- D. SECNAVINST 5236.1B, Contracting for Automatic Data Processing Resources
- E. SECNAVINST 5236.2A, Automatic Data Processing Services Contracts
- F. OPNAVINST 5200.28, Life Cycle Management of Mission Critical Computer Resources for Navy Systems Managed Under the Research, Development, and Acquisition Process

## APPENDIX C

### VENDORS REPLYING TO THE COMMERCE BUSINESS DAILY ADVERTISEMENT

Dataware  
(718)447-4911  
30 Bay Street  
Staten Island, NY  
10301

W J Schaefer Assoc., Inc  
(407)723-4184  
1333 Gateway Dr., Suite 1025  
Melbourne, FL 32901

Houston Fearless  
(213)605-0755

3M  
(612)733-1110  
3M Center  
St. Paul, MN 55144-1000

Mekel Engineering  
(714)594-5158  
777 S. Penarth Ave  
Walnut, CA 91789-  
3072

Minnow  
Micrographics  
(415)872-1182

National  
Micrographics  
Systems, Inc  
(301)588-3200  
926 Philadelphia Ave  
Silver Spring, MD  
20910-4996

Omni Micrographics  
Services, Inc  
(408)945-9805  
1004 Hanson Court  
Milpitas, CA 95035

Tameran, Inc  
(216)349-7100  
30340 Solon  
Industrial Pkwy  
Solon, OH 44139

Visidyne  
(617)273-2820  
10 Corporate Place  
South Bedford Street  
Burlington, MA 01803

## APPENDIX D

### Library Systems Expert Opinion Survey

The purpose of this survey is to collect information to assist in evaluating the importance of each of the benefit factors listed below, in an "ideal" library information system.

Please take a few minutes (five to ten) to provide your view of the importance of each of the following benefit factors.

Rank each benefit factor on a scale of 0 to 10, where 0 means "of no value or benefit" and 10 represents "of the highest value or benefit."

Please identify the systems you most frequently use (i.e., DIALOG, DROLS, RLIN, etc.)

---

- \_\_\_ Acceptance of the system. How the staff views the system, i.e., whether or not the staff believes that the system is useful.
- \_\_\_ Accessibility of information. Speed of access to citations and to the actual information sought.
- \_\_\_ Accountability. Your ability to account for the information in the system.
- \_\_\_ Availability. Access to the system on demand, with little or no waiting to get into the system.
- \_\_\_ Connectivity. The ability to transfer or share information between different systems.
- \_\_\_ Expandability. The ability to add new features and capabilities to the system.
- \_\_\_ Flexibility. The ability for the system to be easily changed or modified to meet new requirements.
- \_\_\_ Maintainability. The ability to easily keep the system "up" and in good operating condition.
- \_\_\_ Mature technology. Having a well established technology with well known procedures.
- \_\_\_ Obsolescence. The degree to which a system is technologically "out-of-date".
- \_\_\_ Productivity. The effectiveness of the system in helping you and other staff to get your jobs done.
- \_\_\_ Quality of searches. The usefulness of the system in helping you to locate the information you are seeking.
- \_\_\_ Reliability. The confidence that you have in the system.
- \_\_\_ Security. The ability to control confidential or classified information.
- \_\_\_ Staff morale. Whether or not using the system adds to or detracts from morale.
- \_\_\_ User friendliness. Ease of use of the system (i.e., It provides enough information about what you can do and how to do it, and has sufficient online "help" available.)

Please identify and weigh any other factors you deem important on the back of this form.

## LIST OF REFERENCES

- Anamet Laboratories, Automated Information Management Systems, Anamet Laboratories, Hayward, California, 1988.
- Blair, D. C., and Maron, M.E., "An Evaluation of Retrieval Effectiveness For a Full-text Document-retrieval System", Communications of the ACM, 28(3):289-99, 1985.
- Bonner, P., "CD-ROM Power: Knowledge in Hand", PC/Computing, February, pp. 64-75, 1990.
- Bromley, K. "Frequency Compression and Expansion Using an Electrooptical Processor", United States Patent 4,236,043, November 25, 1980.
- Buddine, L., and Young, E., The Brady Guide to CD-ROM, Prentice Hall Press, 1987.
- Burnacz, J. G., "Digital Technology Transforms The Reader-Printer", Today's Office, v.25, pp. 12-16, September 1990.
- Bush, V., "As We May Think," Atlantic Monthly, July 1945.
- Buttler, C., "Small Wonder in Delaware", Inform, v.4, pp. 20-23, July/August 1990.
- Colvin, G., "Database Retrieval and Indexing", CD-ROM Volume Two Optical Publishing, Ropiequet, S., ed., Microsoft Press, 1987.
- Desmarais, N., "CD-I vs. DVI", CD-ROM Librarian, pp. 24-28 March 1989.
- Eaton, N. L., McDonald, L., Salue, M., CD-ROM and Other Optical Information Systems: Implementation Issues for Libraries, Oryx Press, 1989.
- Haga, W. J., and Lang, R. G., Economic Analysis Procedures for ADP, Naval Postgraduate School, Monterey, California, 1991.
- International Standards Organization (ISO), Geneva, Switzerland, Letter to Robert W. Clipper, Subject: ISO International Standards for Optical Computer Disc, 22 January 1991.

Interview between Mr. Rick Burrus, Terminal Data Corporation, Moorpark, California, and the authors, 3 January 1991.

Interview between Mr. Jim W. Douglass, W. J. Schafer Associates, Inc., Melbourne Florida, and the authors, 9 January 1991.

Interview between Mr. Ken Hallam, Chairman, ANSI X3B11 Committee, San Juan Capistrano, California, and Robert W. Clipper, 17 January 1991.

Jones, J., "DTIC'S TR CD-ROM Prototype", The Military Librarian, pp. 4-5, October, 1990.

Kaebnick, G. E., "EDMICS Goes to Work," Inform, v.4, pp. 15-18, October 1990.

Lambert, S., and Ropiequet, S. (Eds.), CD-ROM: The New Papyrus, Microsoft Press, 1986.

Lind, D. L., Optical Laser Technology, Specifically CD-ROM, and Its Application to the Storage and Retrieval of Information, Master's Thesis, Naval Postgraduate School, Monterey, California, June 1987.

Lingvai, J. R., "U.S. Army's Personnel Electronic Records Management System (PERMS)", paper presented at the Association of Records Managers and Administrators, Inc., Annual Conference, 35th, San Francisco, California, 5 November 1990.

Manns, B. and Swora, T. "Books to Bits: Digital Imaging at the Library of Congress", Journal of Information and Image Management, pp. 27-32, October 1986.

Mansfield, E., Microeconomics, Theory and Applications, 4th ed., pp. 480-481, W. W. Norton and Company, 1982.

Mekel Engineering, Inc., Technical Data Reduction Sheet, 31 May 1989.

Meridian Data, Inc., CD Publisher Software 5.0 System Manual, p. 48, 1990.

Mueller, R. "Desktop Scanners Come of Age", PC Publishing, pp. 4-25, February 1988.

Naisbitt, J., Megatrends, Warner Books Inc., 1982.

Naval Undersea Center, Report NUC-TR-529, The Bureau of Naval Personnel's Microfiche Image Transmission System: A State-of-the-art Microfacsimile System, by D. Endicot, D. Solarek, and K. Drake, December 1976.

Navy Publications and Printing Service, Washington, D.C., Navy All Ships and Station Directives economic Analysis and Decision Paper, prepared by Integrated Microcomputer Systems, Inc., pp.9-16, January 1990.

Rome Air Development Center, Report RADC-TR-76-150, Microfiche Scanner and Remote Display System, by R. A. Boticelli, D. J. Walker, and G. N. Wallmark, May 1976.

Rome Air Development Center, Report RADC-TR-71-51, System Design Study for Human Readable Machine Readable (HRMR) Information Processor, by W. L. Hayes, H. Ristlaid, and N. S. Szabo, March 1971.

Ropiequet, S., Einberger, J., and Zoellick, B. (Eds.) Optical Publishing: A Practical Approach to Developing CD-ROM Applications, Microsoft Press, 1987.

Rothchild Consultants, "Docucon Offers Conversion Services," Optical Memory News, Rothchild Consultants, July/August 1989.

Saffaday, W., Computer-Output Microfilm, Its Library Applications, American Library Association, 1978.

Sanders, D. H., Computers Today, McGraw Hill, Inc., 1987.

Stanton, T., Burns, D., and Venit, S. "Page-To-Disk Technology: Nine State-of-the-Art Scanners", PC Magazine, pp. 128-177, September 30, 1986.

Taylor, R. R. Conversion of Hard-Copy Documents to Digital Format Utilizing Optical Scanners and Optical Storage Media, Master's Thesis, Naval Postgraduate School, Monterey, California, March 1989.

Telephone conversation between Mr. Don Black, Office of Technical Assistance, General Services Administration, Alexandria, Virginia, and Robert W. Clipper, 28 February 1991.

Telephone conversation between Mr. Ken Caldwell, West Coast Information Systems, Walnut Creek, California, and J. D. Fowler, 20 February, 1991.



Telephone conversation between Mr. Kenneth Hallam, Chairman of ANSI X3B11 Committee, ENDL Associates, San Juan Capistrano, California, and Robert W. Clipper, 20 November 1990.

Telephone conversation between Mr. Dave Kyle, U.S. Navy, Engineering Data Management Information and Control System (EDMICS) project manager, Alexandria, Virginia, and Robert W. Clipper, 27 February 1991.

Telephone conversation between Mr. Basil Manns, Automated Systems Office, The Library of Congress, and Robert W. Clipper, 9 November 1990.

Toffler, A., The Third Wave, Bantam Books, 1981.

U.S. Department of Commerce, Commerce Business Daily, Chicago, Illinois, p. 27, December 6, 1990.

U.S. Department of Commerce, National Institute of Standards, Federal Information Processing Standards, Gaithersburg, Maryland, 1989.

U.S. General Services Administration, Information Resources Management Service, A Guide for Requirements Analysis and Analysis of Alternatives, Washington, District of Columbia, 1990.

U.S. National Archives and Records Administration, Office of Records Administration, Managing Electronic Records, Washington, District of Columbia, 1990.

U.S. Naval Data Automation Command, Economic Analysis Procedures for ADP, NAVDAC PUB 15, Washington, District of Columbia, 1980.

U.S. Naval Postgraduate School, Naval Postgraduate School Catalogue, p. 6, Monterey, California, 1990.

Waegemann, C. P. The Handbook of Optical Memory Systems, Optical Disk Institute, 1989.

Zoellick, B., "Selecting an Approach to Document Retrieval", CD-ROM Volume Two Optical Publishing, Ropiequet, S., ed., Microsoft Press, 1987.

## INITIAL DISTRIBUTION LIST

1. Professor Tung X. Bui 1  
Administrative Sciences Department  
Naval Postgraduate School  
Monterey, CA 93943
2. Lieutenant R. W. Clipper, MSC, USN 2  
Management Information Department  
Naval Hospital  
Pensacola, FL 32506
3. Defense Technical Information Center 2  
Cameron Station  
Alexandria, VA 22304-6145
4. Commander J. D. Fowler, USN 2  
Naval Computer and Telecommunications Station  
Code 50  
Naval Air Station, North Island  
San Diego, CA 92135-5110
5. Professor Barry A. Frew 1  
Administrative Sciences Department  
Naval Postgraduate School  
Monterey, CA 93943
6. Professor William J. Haga 1  
Administrative Sciences Department  
Naval Postgraduate School  
Monterey, CA 93943
7. Mr. Jack Jones 1  
Defense Technical Information Center  
Cameron Station, Building 5  
Alexandria, VA 22304-6145
8. Lieutenant Commander David J. Lind, USN 1  
Naval Computer and Telecommunications Command  
Code N-434  
4401 Massachusetts Ave. N.W.  
Washington, D.C. 20394-5000
9. Commander Rodney Matsushima, SC, USN 1  
Administrative Sciences Department  
Naval Postgraduate School  
Monterey, CA 93943

- |  |   |
|--|---|
| 10. Commanding Officer<br>Naval Health Sciences<br>Education and Training Command<br>Bethesda, MD 20814-5022           | 1 |
| 11. Commanding Officer<br>Naval Medical Data Services Center<br>Bethesda, MD 20814                                     | 1 |
| 12. Library, Code 52<br>Naval Postgraduate School<br>Monterey, CA 93943-5002   | 2 |
| 13. Mr. Joseph W. Price<br>Chief, Science and Technology Division<br>The Library of Congress<br>Washington, D.C. 20540 | 1 |
| 14. Captain R. A. Weaver, MC, USN<br>Naval Aerospace Medical Institute<br>Code 42<br>Pensacola, FL 32506               | 1 |