Image Storage and Retrieval Systems

A New Approach to Records Management

Dr. Marc R. D'Alleyrand

Image Storage and Retrieval Systems

A New Approach to Records Management

Dr. Marc R. D'Alleyrand

Intertext Publications
McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá
Hamburg London Madrid Mexico Milan Montreal
New Delhi Panama Paris São Paulo
Singapore Sidney Tokyo Toronto

•Library of Congress Catalog Card Number 89-83761

10 9 8 7 6 5 4 3 2 1

ISBN 0-07-015231-4

Intertext Publications/Multiscience Press, Inc.
One Lincoln Plaza
New York, NY 10023

McGraw-Hill Book Company
1221 Avenue of the Americas
New York, NY 10020

Composed in Ventura Publisher by Context, Inc.

Introduction

0.1. PURPOSE AND ORGANIZATION OF THIS BOOK

The most far-reaching advances in the recent history of office automation have been in the development of electronic imaging systems. The ability to store and retrieve computer data and images within a single installation is creating new opportunities for improved productivity. It also raises a host of issues that, if not properly resolved, can change these opportunities into prescriptions for failure.

This book is for users and designers who need to deal with issues that are essential to the success of any office automation project. As a guide to the proper use of modern storage and retrieval technologies, it is sophisticated enough to meet the needs of managers and decision makers seeking a broad understanding of the issues involved. At the same time, the book is written with a minimum of technological jargon, so that it can be used by readers with limited technical expertise; therefore, it can be used by line personnel who want to improve their professional skills. Directed to a large audience, including both practitioners and students in the field, it stresses the influence of business considerations on the acceptance of office automation systems, and shows how traditional information-management techniques can be applied to facilitate their implementation.

This book focuses on three interrelated functional responsibilities (see Figure 0.1.):

- The manager, who must understand how to use office technologies within the context of organizational needs and profit goals.
- The systems designer, who must translate organizational goals into a cost-effective design within short-term and long-term constraints.

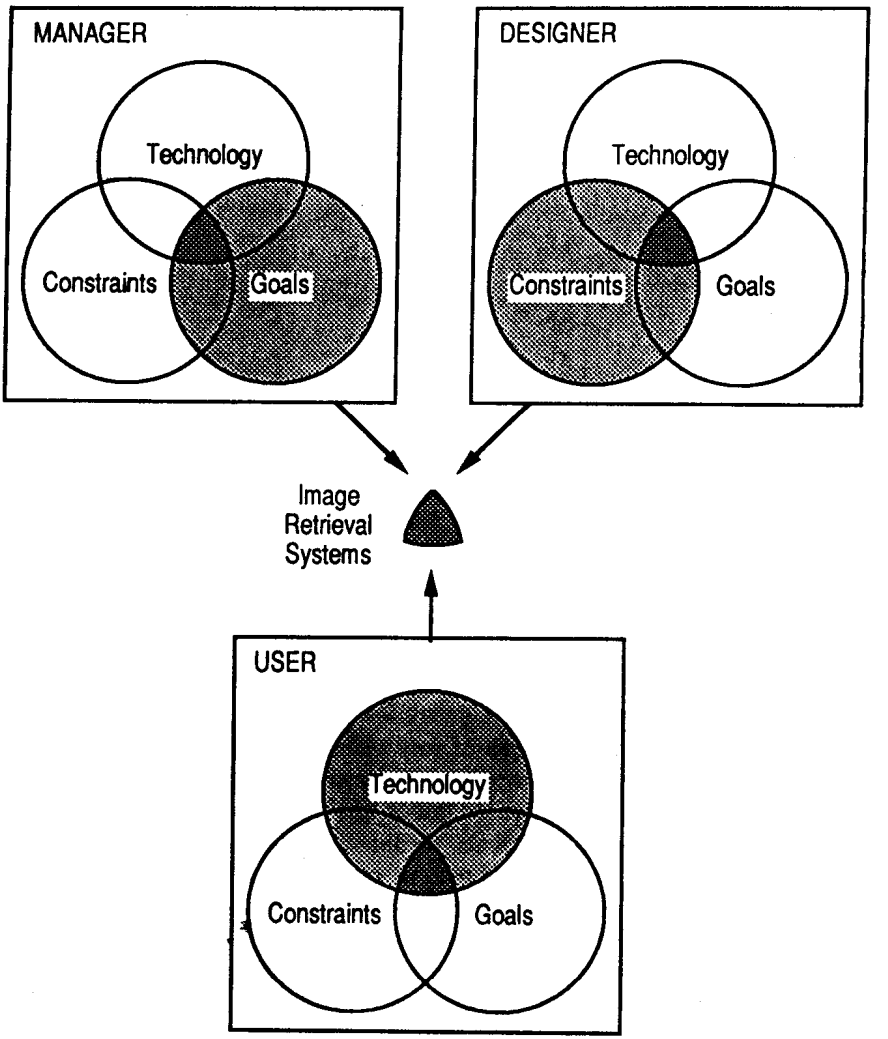

Figure 0.1 The clientele of image retrieval systems.

- The end user, who needs to understand the limitations of the technologies, so that he can specify realistic operating requirements.

In a small operation, the same individual often assumes all three responsibilities. For instance, in small offices, the end users sometimes select and acquire equipment directly and set up the procedures for its use. In most cases, however, different people handle the

three functions. They need to communicate effectively and operate as a team, regardless of differences in their backgrounds and experiences.

It is hoped that a book that considers the needs and perspectives of each function can reconcile their diverging interests and result in a unified approach. However, because no two environments are alike, the ideas contained in these pages can only provide directions, not definite solutions. In the final analysis, the successful application of these concepts to a particular situation remains the reader's responsibility.

To allow quick access to the concepts it discusses, *Image Retrieval Systems* is written in a modular fashion. The introduction places documents and images within the context of office automation in general. The body of the book discusses their handling in the specific area of retrieval systems. It is organized around three main themes: structure, equipment, and management. An inverted index, based on the table of contents, ports this modular structure in a highly efficient manner.

Part A, a primer on information retrieval, reviews the major differences between image and data systems, discussing the application of records management techniques to the design of such systems. Chapter 1, an overview of the information field, discusses the origin of the so-called information explosion and some of the ideas behind the design of retrieval systems. Chapter 2 covers the basic principles of indexing, and Chapter 3 reviews fundamental design trade-offs for imaging systems. Chapter 4 places the issue of imaging systems within the context of common organizational constraints.

Part B is dedicated to studies of the various technologies commonly found in imaging systems, focusing on those that are not usually discussed in the context of data systems or that are not usually considered to be an integral part of the imaging system itself. Chapter 5 provides a general overview of the technological environment as it supports the structure of a retrieval system. Chapter 6 covers the various data capture technologies, with a discussion of the problem of validation; a section in this chapter is dedicated to the emerging smart card technologies. Chapter 7 reviews image capture, with specific mention of the problems of legibility and quality control as they apply to microfilming and, more generally, to optical disk systems. Chapter 8 deals with the processing function, with special emphasis on the implication of task and function system designs. Chapter 9 deals with the technologies used for image retrieval, transmission, and display. Chapter 10 applies the concepts discussed to the design of optical disk systems.

Part C provides a blueprint for implementation. Chapter 11 deals with the assessment of realistic requirements for the retrieval system. Chapter 12 treats design issues, and Chapter 13 considers implementation strategies. Chapter 14 reviews some of the operational issues that must be dealt with throughout the life of a retrieval system.

0.2 OFFICE AUTOMATION IN CONTEXT

Office automation is a generic term for the application of data processing technologies to the performance of office tasks. The use of the digital technologies allows organizations to handle information flow in a much faster way than they could in a manual environment. Two common examples are word processors, designed to speed up the generation of correspondence, and facsimile machines, which facilitate interoffice communication. In theory, office automation, by accelerating information flow, allows better controls and speedier decision making and generates an overall increase in productivity.

In practice, however, results are often not impressive. The ease with which documents and reports can be generated can also create operational bottlenecks. Electronic mail systems can be burdened by "electronic junk mail." In many instances, savings do not materialize or are offset by new costs. In other cases, new systems are operationally cumbersome, cannot respond to changes in the business environment, or become quickly obsolete. These unanticipated situations often have a variety of causes. Most of the time, they reflect an inability of system users and designers to articulate their needs, requirements, and constraints to each other. Without such articulation and the ability to understand the problems of office operations, office automation can only mechanize preexisting conditions rather than provide reasonable solutions.

0.2.1. Office Automation the Wrong Way

Inefficient office automation means a loss of opportunities for overall productivity increases, often caused by a misunderstanding of the difference between tasks, such as filing and typing, and functions, such as order processing. Furthermore, improvements in isolated routine activities do not necessarily result in overall operational gains. Without optimization of the information transfer process over the entire information flow, gains in one task might mean lower per-

formance in other areas. Under these conditions, net productivity gains remain low. This lack of optimization of processes was seldom taken into consideration in early studies of the limited success of office automation. Rather, these studies concluded that low productivity was the result of insufficient investment in equipment.

This opinion was strongly affirmed in the classic "study on managerial and professional productivity" conducted in the late 1970s by the consulting firm of Booz-Allen & Hamilton. That survey recorded between 3,000 and 8,000 time samples in each of 15 major industries. The results were subject to more than 100 man-months of analysis and evaluation. The conclusion was that to better meet their information needs, especially those essential to profit improvement, organizations needed to support their "knowledge workers," who were found to spend on the average close to 50% of their time in meetings and telephone calls (see Figure 0.2) That view, derived from a wide array of credible data, mandated heavy capital investment in office automation. The argument was simple and apparently logical: If automation of office tasks allowed managers and knowl-

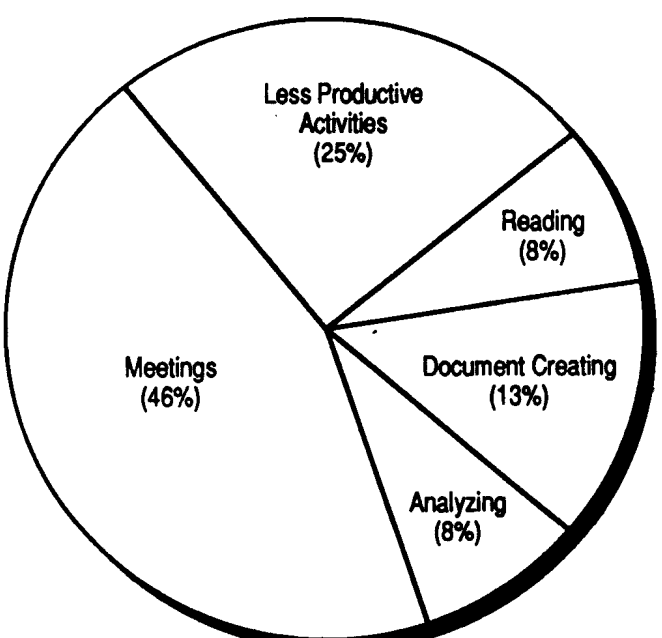

Figure 0.2 Allocation of work time by knowledge workers (source Booz-Allen & Hamilton).

edge workers to make better use of their time, they would be able to increase their productivity. What made this conclusion difficult to accept was that it seemed to repeat prior recommendations of the information industry. Early concepts, such as electronic data processing, and later ones, like management information systems, had also promised that the use of computer-based systems would reduce overall costs by improving the transfer of information. These promises had often been left unfulfilled.

The novelty of the Booz-Allen study was not so much the argument that computerized systems would improve productivity, but the fact that systems were now adapted to the office environment. Furthermore, the postulate that capital investments in technologies could be as effective in the office as in the factory assumed that factory and office environments offered similar opportunities for automation. What this postulate ignored was that operating conditions in the two environments were often entirely different, and therefore could not be analyzed in the same terms. In particular, it did not recognize that the office–factory parallel could apply only to repetitive tasks, performed by clerical rather than by knowledge workers. A more seriously flawed conclusion was the implication that lessons learned from the survey could be duplicated. This idea is difficult to accept, because the operating conditions of future systems could not be the same as those reported in the study. In retrospect, the study reaffirms the intuitive feeling that an increase in productivity requires a thorough understanding and proper management of office technology implementation.

The fact is that office automation projects are not always able to deliver their expected benefits. Often low levels of performance are only the last in a series of frustrations that include cost overruns and delays in implementation. The success of office automation depends on factors other than the brute infusion of capital investment, and it is the purpose of this work to provide guidance about "a better way."

0.2.2. Maximizing Productivity Through Automation

Finding that "better way" is crucial. Automated systems are now familiar features in many offices: stand-alone or officewide word processors, microforms, sophisticated electronic mail, and optical disks. Even smaller offices without internal automation depend on new technologies for their communication with the outside world through rented, sophisticated telephone systems. Consequently, no modern

organization can deny that office automation is a necessary ingredient in business life. And because information retrieval is an essential component of office activity, information retrieval systems must play a part in the productivity issue.

The importance of information retrieval is underlined by the industry projections concerning market growth. When IBM announced its entry into the image management field, the periodical *MIS Week* (*MIS Week*; July 4, 1988, p. 10) estimated that the imaging systems market would be close to $3 billion by 1991, with the largest contribution to come from small and medium-size systems (see Figures 0.3 and 0.4). With such numbers, even if projections are optimistic, there is no doubt that office automation is here to stay. The sheer pressure

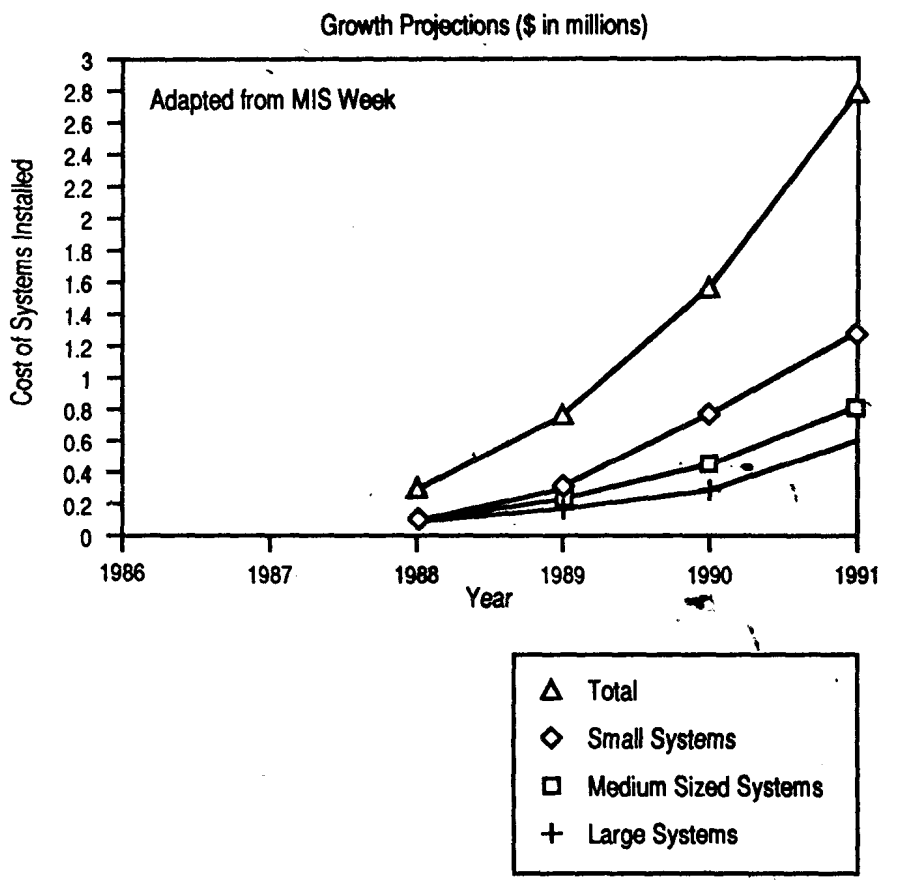

Figure 0.3 Imaging systems market growth projections in $ million (source *MISWeek*)

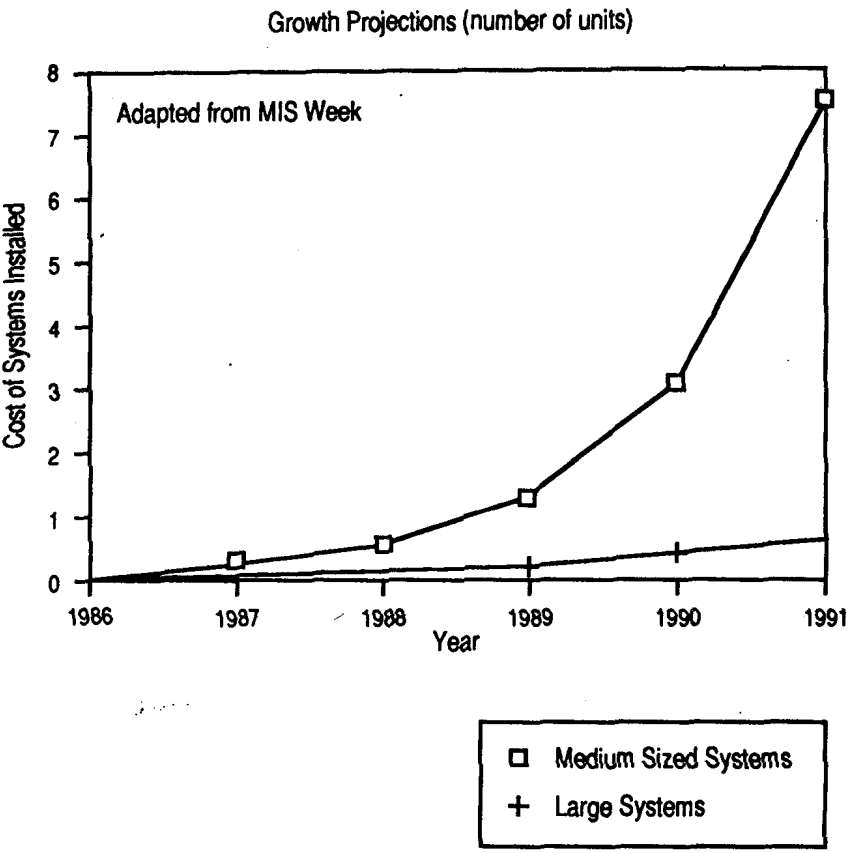

Figure 0.4 Imaging systems market growth projections in 100 units (source *MISWeek*).

to remain competitive will force even the most conservative organizations to adopt office technologies. Once the implementation process has begun, it becomes irreversible, even if increases in efficiency and profits are not as high as expected.

Office automation forces permanent changes in organizational life. Regardless of the reasons for adopting it, after automation there is no return to prior ways of conducting business. Consequently, the operating costs associated with the system become a necessary cost of doing business, with direct impact on the profitability of the organization. This reality shifts the fundamental decision issues from "Should one automate?" to "How to automate and when?" in a way that ensures minimum risks for maximum returns.

The long-term financial commitment necessary for systems automation requires that these systems adapt to largely unpredictable changes in operating conditions, changes sometimes as profound as organizational restructuring, acquisition, and merger. The ability to respond will be influenced by concrete factors, such as adequate hardware and software, as well as by a combination of less tangible elements, such as procedures, people, and training.

Flexibility and an ability to respond to change are difficult to obtain in a totally automated environment, which, in order to operate efficiently, needs to be structured. The adoption of some practices derived from the operation of systems provides the needed flexibility. Because, by nature, automated procedures cannot accommodate all situations, some provisions must be made to allow for unstructured procedures to handle the unanticipated. As a result, the most successful installations are those in which the most routine operations are automated, but in which human intervention is possible whenever judgment is necessary or when an unexpected condition occurs.

0.2.3. Avoiding Pitfalls in Design

To maximize the potential for productivity gains, systems engineers and planners must ensure that the level of sophistication of the system being designed is adapted to the complexity of the tasks to be performed. If the system is too sophisticated, it will be underutilized. If it is not sophisticated enough, it will not be able to perform adequately. In either case, expected productivity gains are unlikely to materialize.

The need to adapt system sophistication to the level of tasks leads to the question of whether only trained personnel should oversee the implementation and use of automated systems. Word processing systems provide a good example of such an issue: Because equipment is designed to appeal to the largest clientele, there is a tendency to include in such systems more capabilities than are necessary in most document-preparation situations. The great number of features might be useful to the trained secretary but will often exceed the average user's requirements, thus reducing the cost effectiveness of the system and potential productivity gains. The advantage of using trained personnel would likely yield technically better results; in the word processing example, trained personnel would produce perfectly typed documents. The drawback would be an overall reduction in flexibility, as technicians valued for their abilities to handle the mechanical aspects of a task might not be capable of dealing satisfacto-

rily with its purpose or context, thus limiting overall cost effectiveness.

Because the implementation of any office automation system requires the establishment of a structured environment, it is essential to evaluate situations in which structuring of some tasks might result in the destructuring of others. The adoption of a systematic decision-making process during the study and design phase of an automation project will achieve that result. Such a process will require the following (see Figure 0.5):

1. Identification of the structured tasks through a clear definition of organizational objectives and an understanding of the level of performance necessary to meet them.
2. Recognition and assessment of potentials and limitations of the resources and technologies that will be used to meet the overall objectives.
3. Determination of the critical trade-offs between tasks that can be automated at an acceptable cost and those for which automation is not to be considered.
4. Establishment of acceptable interfaces between structured and nonstructured tasks in the automated environment.

The last step is essential to the success of automation. Without it, the system will not function properly, as it will be unable to pass information among its structured and nonstructured components. It takes on particular importance for document retrieval, because unstructured information, carried by images, represents the bulk of the information necessary to the operation of an organization.

0.3. DOCUMENT RETRIEVAL AND RECORDS MANAGEMENT: AN OVERVIEW

Records represent physical evidence of some kind of activities. As such, they are essential to the life of an organization, allowing it to verify progress and maintain controls. Depending on the nature of the activity, records can consist of paper documents, models, parts of machinery, or samples. Documents can be defined as physical groupings of records into meaningful entities. Consequently, document retrieval implies records retrieval. Records management is a discipline developed initially to ensure cost-effective retrieval in manual systems. Office automation technologies by themselves have not changed the principle, only the method. For instance, locating docu-

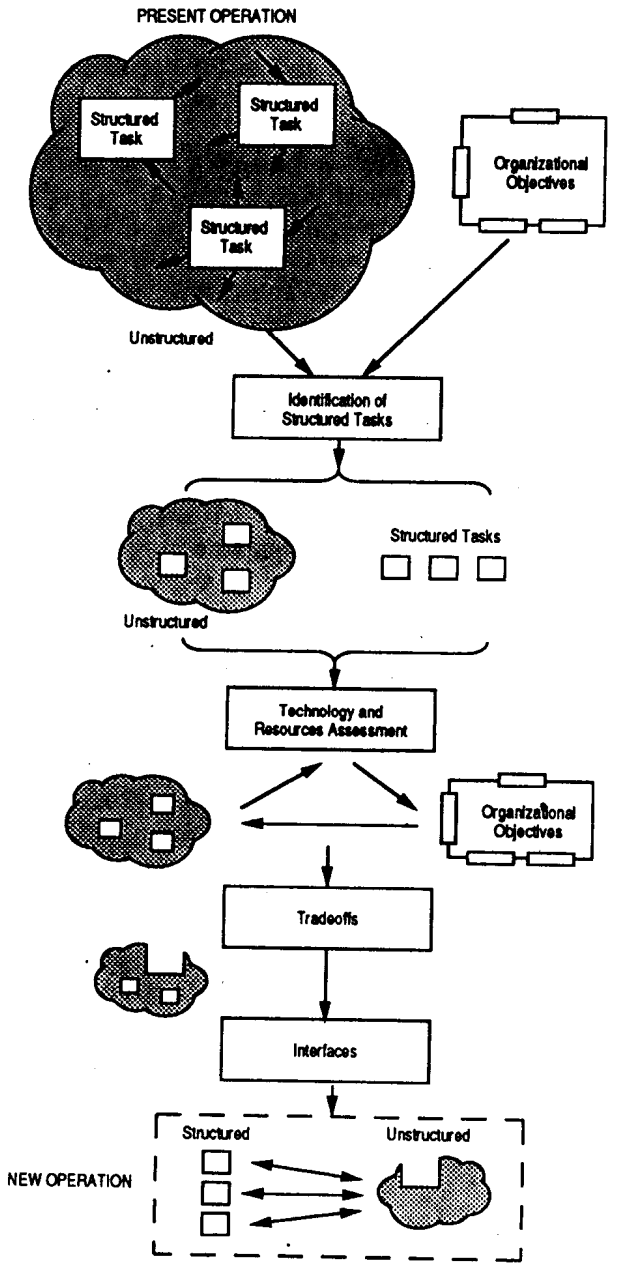

Figure 0.5 Structuring for office automation.

ments in large files is usually faster when the search is done with a computer; however, documents still need to be identified before being retrieved. Because the body of knowledge for document retrieval is well established in records management, it is logical to extend it to the field of automation.

In its simplest form, a records management program focuses only on the storage of documents, without concern for a justification of their existence or for their retrievability. In its most sophisticated form, it covers the reasons, the processes, and the means necessary to the documents' existence and use. To achieve that level of sophistication, which is essential to the continued success of an automated retrieval system, a records management program must cover the entire life span of a document, from its generation to its destruction. One of the major difficulties in setting up such a program is the need to accommodate in one system the handling of documents as discrete physical entities, as well as the handling of their contents, and not necessarily those attached to particular documents. For instance, a piece of correspondence might be related to a subject not mentioned explicitly in the document; its handling as an isolated record would likely result in an incomplete file on that subject. In a manual environment, the physical handling of documents (filing and retrieving) is essentially independent from the processing of their contents (identifying their subject matter). In contrast, the use of electronic imaging in automated retrieval systems requires that these two aspects be linked.

The many components of records management — document capture, indexing, storage, retrieval, and dissemination — are discussed throughout this book as they apply to image systems. For the moment, it is sufficient to understand that, because the purpose of records management is to optimize the use of resources necessary to the retrieval of records, its applicability to cost-efficient document systems is essential. In particular, the techniques of records management are helpful in an early evaluation of opportunities for cost reduction and potential productivity increases.

0.3.1. Cost Considerations

Because retrieval is the last activity necessary to gain access to a document, its cost effectiveness requires the optimization of all preceding steps. Cost effectiveness of transactional activities, such as document capture, contents identification, and actual retrieval, can be analyzed easily because their costs are incurred only when docu-

ments are needed or used. In contrast, the cost effectiveness of inactive storage is more difficult to establish because the cost of storage is incurred whether the documents are used or not (see Figure 0.6).

The overall justification of the costs associated with the design, installation, and operation of document storage and retrieval systems is based on two considerations: They must ensure the timely availability of records, and the costs must be lower than the penalty for not being able to produce the documents within the required time frame.

Regardless of their formats, and excluding records for which laws and regulations set up specific retention periods, there is no systematic method of predicting when records, once used, will be needed again. As a result, any cost justification for a storage system, manual or automated, should include an assessment of the length of time beyond which the costs of continued retention of given documents will exceed their probable value to the organization. That assessment will often be a matter of judgment, directly related to two business

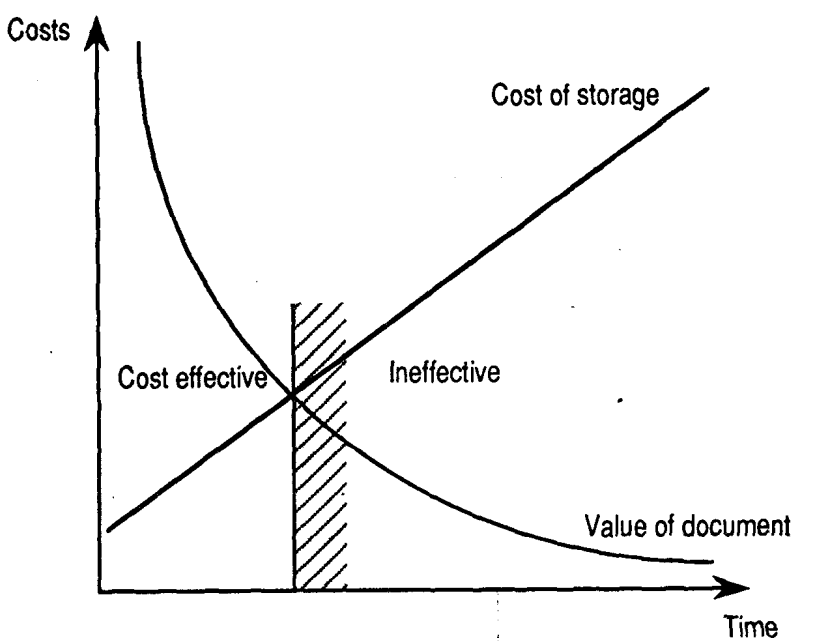

Figure 0.6 Cost effectiveness of document storage.

considerations: the exposure of not having the records at all, and the cost of reconstructing them on demand from secondary sources. This last consideration has been instrumental in the development of many microfilm systems (see Chapter 7), in which the occasional cost of copying from microfilmed documents was cheaper than the cost of the systematic storage of original paper records. It must be understood that the actual assessment of the costs and benefits of record retention depends on the nature of the documents and on the probability of future need. Because records management can provide the tools necessary to that assessment, an understanding of the foundations of that discipline will be invaluable in evaluating the desirability of automating document retrieval.

The most common application of records management techniques to cost analysis is in the area of storage and retrieval. The cost of handling records, which is a labor-intensive operation in a manual office environment, can account for as much as 20% of an organization's overhead. When space is at a premium, annual storage costs alone have been estimated to be more than $300 per file drawer (see Table 0.1).

Table 0.1 Annual Cost of Storing Records

NEEDS		
Space: 5-drawer file cabinet, including aisle space; 6 sq. ft rental $20/sq. ft.		
Personnel:		
Clerical: 1 full-time equivalent for every 12 cabinets; Anual salary $18,000		
Supervision: 0% of clerical costs		
Equipment: File cabinet depreciated over 7 years; purchase price		
Supplies, etc.: 10% of overall costs		
COSTS		
Space	6 sq. ft. at $20/sq. ft	$ 120
Clerical	1/12 of $18,000	1,500
Supervision		150
Equipment	1/7 of $350	50
Total excluding supplies		1,720
Supplies		180
Total per 5-drawer cabinet		1,900
Cost per drawer		$ 380