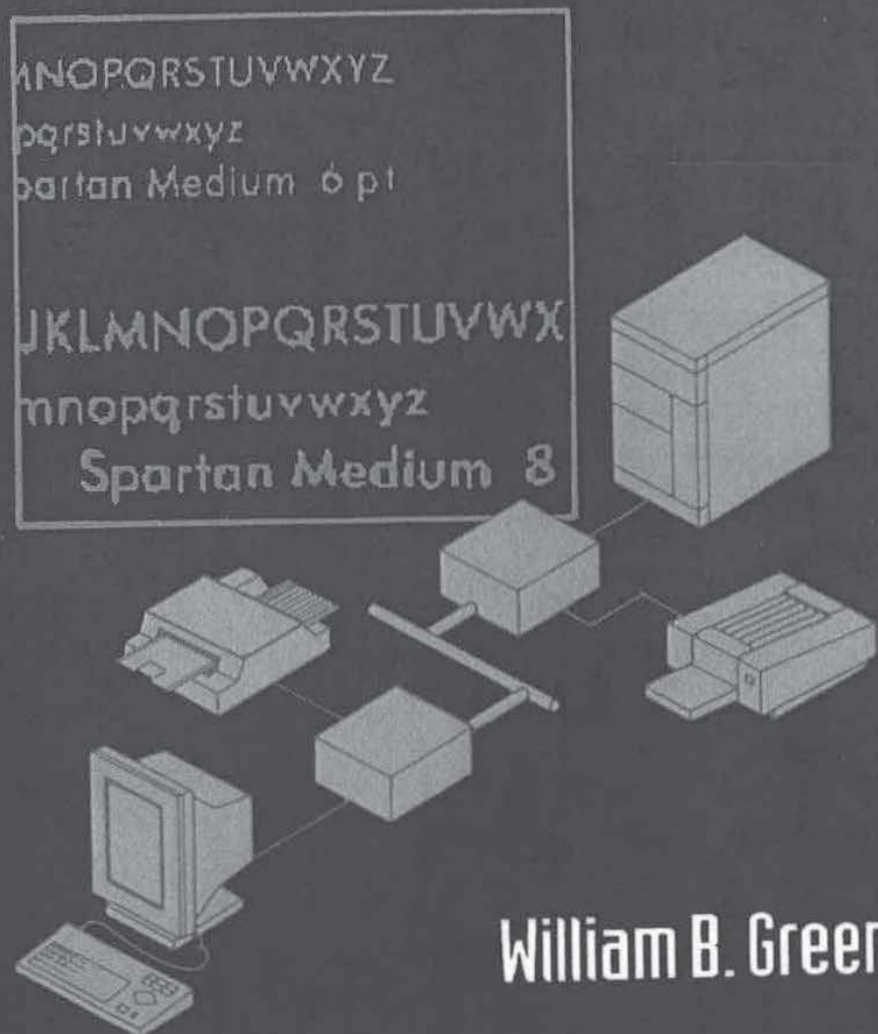


# Introduction to ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS



**Introduction to  
ELECTRONIC DOCUMENT  
MANAGEMENT SYSTEMS**

# Introduction to **ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS**

**William B. Green**

Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California



ACADEMIC PRESS, INC.  
Harcourt Brace Jovanovich, Publishers  
Boston San Diego New York  
London Sydney Tokyo Toronto

Copyright © 1993 by Academic Press, Inc.  
All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Cover photo: High speed scanning at Empire Blue Cross and Blue Shield  
courtesy of Sigma Imaging Systems, Inc.

ACADEMIC PRESS, INC.  
1250 Sixth Avenue, San Diego, CA 92101-4311

*United Kingdom edition published by*  
ACADEMIC PRESS LIMITED  
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Green, William B.

Introduction to electronic document management systems / William

B. Green.

p. cm.

Includes bibliographical references and index.

ISBN 0-12-412102-0

1. Optical scanners. 2. Optical storage devices. I. Title.

TK7882.S3G73 1993

651.5'0285'66—dc20

92-35002

CIP

Printed in the United States of America

93 94 95 96 EB 9 8 7 6 5 4 3 2 1



*This one is dedicated  
to  
Micah and Derek  
and the future they will inherit...*

Inc.

roduced or  
is, electronic  
cording, or  
ystem, without  
ir.

Empire Blue Cross and Blue Shield  
c.

101-4311

lication Data

ent management systems / William

ces and index.

tical storage devices. I. Title.

92-35002  
CIP

5 4 3 2 1

## Contents

Preface .....	ix
---------------	----

### Section 1 INTRODUCTION

CHAPTER 1	
Introduction .....	1
CHAPTER 2	
Image Acquisition .....	25
CHAPTER 3	
Image Display .....	65
CHAPTER 4	
Board Level Components .....	75
CHAPTER 5	
Image Storage on Optical Disk Based Systems .....	87
CHAPTER 6	
System Design Considerations .....	107
CHAPTER 7	
Examples of Electronic Document Management Systems .....	127

Section 2  
ADVANCED TOPICS

CHAPTER 8	
Digital Image Processing Techniques .....	155
CHAPTER 9	
Gray Scale and Color Representation .....	177
CHAPTER 10	
Image Compression .....	193
CHAPTER 11	
Character recognition Technology .....	219
CHAPTER 12	
Bar Code Technology .....	233
APPENDIX	
Sources of Additional Information .....	243
INDEX .....	247

Computers have been used in exploration programs, first digitizing the *Ranger* and *Sun* images, extended to support image representation, specialized computer programs routinely used throughout the field every day.

As in any rapidly changing field, reference sources for the latest developments of the technology of electronic documents. The first section of the book and software used in imaging technology, including technical detail.

The second section of the book covers specific topics, including image representation, image code technology, and other topics, including a summary of technical conferences.

## Contents

### Section 2 ENHANCED TOPICS

.....	155
.....	177
.....	193
.....	219
.....	233
.....	243
.....	247

## Preface

Computers have been used to process pictures since the early days of the space exploration program in the 1960s when Jet Propulsion Laboratory (JPL) engineers first digitized and then enhanced images of the moon returned by the *Ranger* and *Surveyor* spacecraft. In recent years, the technology has been extended to support computer based management of document material, using image representations of document material and a variety of general purpose and specialized computer equipment and peripherals. Imaging technology is now routinely used throughout the world to process hundreds of thousands of documents every day.

As in any rapidly emerging technical field, there is a shortage of standard reference sources for those interested in becoming conversant with the basic elements of the technology. This text strives to provide a basic introduction to electronic document management systems. The text is written in two sections. The first section provides a general introduction to the methodology, equipment and software used to process document material using computer based electronic imaging technology. It is written for the general reader, with a minimum of technical detail.

The second section of the text provides a more detailed introduction to specific topics, including digital image enhancement, gray scale and color image representation, image compression, character recognition technology, and bar code technology. This section provides an in-depth treatment of each of these topics, including occasional use of mathematical equations. An appendix provides a summary of additional sources of information, and a summary of useful technical conferences and symposia.



This text is intended for use by individuals in industry, government, the military, universities, financial institutions, and other organizations faced with large scale information management problems who are considering using electronic document management technology to accommodate increasing volumes of information and the need to reduce response times in information management systems. It can also be used in an upper division or graduate level course in information management or computer science as the basis for a one semester course in document management technology, possibly supplemented with one or more of the references provided in each chapter.

Just a few years ago, those of us working with electronic document management systems as integrators or vendors could walk the aisles of the Association for Information and Image Management (AIIM) show and visit with almost everyone else involved in the technology. Today, the AIIM show has grown into one of the largest trade shows in the country, with large and diverse exhibits by all the major computer industry corporations. Hopefully this book will help those new to this dynamic and expanding technology understand the many issues they will encounter while trying to apply that technology to their activities.

I have been fortunate to participate in the growth of this field based on the circumstance of being in the right place at the right time. At JPL during the 1970s, I was fortunate to learn the technology of digital image processing while participating in the unmanned exploration of the solar system. In the 1980s, after arriving at System Development Corporation (later acquired by Burroughs and then merged into Unisys), I was able to apply some of the image processing techniques from JPL as part of the implementation of the Automated Trademark System at the U. S. Patent and Trademark Office under government contract. At Unisys we were involved in bidding two components of the Automated Patent System and we then bid and won the contract for the Optical Digital Image Storage System (ODISS) at the National Archives and Records Administration (NARA). Through ODISS, I met Bill Hooton, one of the pioneers in electronic document management, who effectively articulated his views of the future potential of this technology. At Unisys, I was fortunate to work with Joe Hughes, who is another of the pioneers of this technology, and with Dave Vandaveer, who successfully managed the ODISS project to its completion.

At Terminal Data Corporation from 1987 through 1990, I was fortunate to lead design teams developing some of the first commercially available paper and film scanners. At TDC, I had the opportunity to work with a wide range of system integrators, major equipment manufacturers, and value added resellers, all of them struggling in the late 1980s to deliver economically viable systems to customers willing to take the risks of investing in new technology. At TDC, I was privileged to work with David Krar Hutton, who supported our design efforts as President and CEO during some turbulent times in the industry, with Connie Moore, who provided strategic market insight to guide our engineering

efforts,  
as we n  
lines fa  
Rothbar  
and wit  
TDC un

No o  
assistan  
illustrat  
resentat  
Inc., Si  
Intergra  
Eastma  
PRC In  
Corpor  
Enginee  
System:  
the AIIM  
ous mat

I aga  
folks v  
Cheryl  
third bc  
uing so  
underst  
crazy th  
ogy tha  
our two

Fina  
in his f  
places



individuals in industry, government, the military, and other organizations faced with large problems who are considering using electronic technology to accommodate increasing volumes of data and response times in information management. This is an upper division or graduate level course in computer science as the basis for a one semester course in technology, possibly supplemented with one or more chapters.

Working with electronic document management systems could walk the aisles of the Association for Information Management (AIIM) show and visit with almost any technology. Today, the AIIM show has grown into a country, with large and diverse exhibits by corporations. Hopefully this book will help expanding technology understand the many ways to apply that technology to their activities. It is in the growth of this field based on the circle at the right time. At JPL during the 1970s, the study of digital image processing while participating in the solar system. In the 1980s, after arrival (later acquired by Burroughs and then applied) some of the image processing technology implementation of the Automated Trademark Office under government contract. At the two components of the Automated Patent Office contract for the Optical Digital Image Archives and Records Administration. Bill Hooton, one of the pioneers in electronic technology articulated his views of the future. At Unisys, I was fortunate to work with Joe and engineers of this technology, and with Dave and the ODISS project to its completion.

From 1987 through 1990, I was fortunate to find the first commercially available paper and opportunity to work with a wide range of system manufacturers, and value added resellers, all to deliver economically viable systems to those investing in new technology. At TDC, I

Krar Hutton, who supported our design during some turbulent times in the industry, with his market insight to guide our engineering

efforts, with Harold Gross, who provided continual support and encouragement as we merged the manufacture of electronic imaging products into production lines famous for high precision micrographics products, with the late Mike Rothbart, who understood the need to transform the company he had founded, and with a great group of engineers who continue to design great products for TDC under Ed Gonser's leadership.

No one writes a book on their own, and this book is no exception. I received assistance from many individuals in the industry and government who provided illustrative material for use and reviewed portions of the text. They include representatives from TDC, Unisys, Photomatrix, FileNet, Seaport Imaging, Xionics Inc., Sigma Imaging Systems Inc., Empire Blue Cross and Blue Shield, Intergraph, Syncrude Canada Ltd., Kofax Image Products, Bell & Howell, Eastman Kodak Company, the National Archives and Records Administration, PRC Inc., Litton Industrial Automation, AIM USA, Fujitsu America Inc., Sony Corporation of America, Cygnet Systems Inc., Ricoh Corporation, Mekel Engineering Inc., Document Technologies Inc., Cincinnati Bell Information Systems, Rapid Technology Corp., and the IEEE. And a special thanks goes to the AIIM Resource Center, who provided invaluable help in tracking down various material and standards.

I again thank my family for continuing support and encouragement—all the folks who appeared on the dedication pages of my first two books—Bob, Cheryl, Rich, Jennifer, Kimberly, Max, Dorothy, Yetta and Nathan. This is my third book, and I managed to marry a wonderful woman who has been a continuing source of strength and good counsel through all of them. Barbara seems to understand that every few years I disappear on nights and weekends to do this crazy thing. Many thanks again, honey. Since this book deals with new technology that will evolve into the future, it seemed appropriate to dedicate this text to our two new grandchildren, whose arrival interrupted progress on the manuscript.

Finally, thanks to our cat Murphy, who has kept me company for many hours in his favorite warm spot next to my Macintosh. Hopefully, we've found all the places where his paw rested on the backspace key for a while...

William B. Green  
Granada Hills, California  
August 1992

## Section 1

# INTRODUCTION



## Introduction

Commercial firms, government agencies, medical services, educational institutions, and other organizations are faced with an explosive increase in the amount of paper that must be handled in order to conduct their operations. These organizations are continually seeking systems that can solve a variety of problems associated with handling large volumes of paper. The problems include the need for large volumes of storage space required to retain the documents, labor intensive operations associated with filing, retrieving, and routing paper documents, and an increase in processing time in transaction oriented operations in an era when competitive pressures and an orientation toward customer service dictate a reduction in processing time.

Micrographic systems, which have been in existence for many years, involve use of film (typically microfilm, microfiche, or aperture cards) for compact long term storage of document material. Micrographic systems incorporate a photographic process that captures an image of every document stored or processed by the particular application. Once the photographic image is created, it generally becomes the primary source of information within the application, and the film product is used to view the document material, route it from point to point within the enterprise, and for long term archival storage and retrieval. Specialized micrographics equipment is used to create (through photographic means), store, retrieve, display, and generate hard copy from filmed versions of source documents. Special film viewers are used to examine document material, and printers can be used to create hard copy paper versions of filmed document images.

Micrographics systems are found in almost every field of human endeavor. Applications include distribution of document material on microfiche cards, archival storage of check images on microfilm in the banking and financial com-



munity, storage of engineering drawings on aperture cards, and the long term storage of medical records, insurance claims forms and other documents on microfilm. A computer based index to film records is often created to aid in locating specific documents on individual physical film records. Computer based systems are available that support query and retrieval of specific document images from the appropriate film roll, microfiche card, or aperture card.

In recent years, electronic document processing systems have emerged that provide alternatives to physical processing of paper or the use of micrographic systems. Electronic document processing systems perform functions similar to the older micrographic systems. In electronic document management systems, however, the process of filming documents is replaced by a process of electronically scanning document material. Electronic document processing systems process documents stored in digital image format that are created by sophisticated electronic scanning equipment. Once the document material is converted into a digital image format, it can be routed, displayed, stored, and retrieved using commercially available computer equipment and software designed for the specific application. Electronic systems have also been used to replace micrographics systems, and in some cases a phased transition has occurred that has involved scanning of the document database originally stored on film and conversion of that database to electronic format using specially designed film scanning equipment.

Most electronic document processing systems include the following types of hardware subsystems:

- i) Scanners that scan document material and convert it to a digital image format for storage and processing by computer compatible equipment
- ii) Storage systems that are used to store large numbers of scanned images
- iii) Workstations equipped with high resolution displays for viewing the scanned images
- iv) Communications subsystems that transfer images between elements of the system
- v) Hard copy output devices (printers) used to provide copies of the scanned images on paper
- vi) A variety of computer peripheral devices, including magnetic and optical disk subsystems and magnetic tape subsystems

This chapter begins with a description of the conversion of document material into digital image format, which can be processed, manipulated, and displayed using computer based equipment. A more complete description of the hardware elements is presented, followed by a discussion of the advantages of using an electronic system to process documents. A breakdown of the major categories of electronic document processing systems is presented, and the organization of the text is then described.

## HOW

Document management can be managed in many types of digital image formats. This section describes digital representation of documents and the terms that are used.

## Definitions

In this text, *page* refers to a single page of paper. A *bit* represents either one or two bits depending on whether the image is of one or more sheets of paper in a document or in applications or in

## Bit Mapped

Selection of a bit map is one or two digits of paper. A *bit* represents a picture element stored as either a bit or a byte referred to as a *bit*.

Figure 1.1 shows a variety of bit maps which shows a bit map was created by a bit mapped representation described in detail.

Figure 1.3 shows a bit map. In the enlargement in Figure 1.2 are higher resolution individual black and white original page.

Each sampled point, or pixel for

## HOW COMPUTERS PROCESS DOCUMENTS

Document material must be converted to a computer compatible format before it can be managed by an electronic document processing system. There are several types of digital formats commonly used in electronic document processing. This section describes the options available for conversion of document material into digital representations, indicates the advantages and disadvantages of each, and it describes the type of format used in the electronic document management systems that are the focus of this text.

### Definitions

In this text, *page* is the term used to indicate a single piece of paper in the original paper file. A *document* is a collection of one or more pages. A *page* can generate either one or two images to be stored and processed by the digital system, depending on whether or not it is necessary to handle both sides of each original sheet of paper in the particular application. A *file* is used to describe a collection of one or more documents. A *file* is often also referred to as a *folder* in some applications or in the literature.

### Bit Mapped Image Format

Selection of a bit mapped image format means that each page is converted to one or two digital images, each of which represents one side of the original sheet of paper. A *bit mapped* image is a two dimensional array of points that represents a picture of the original page. Each point in the two dimensional array is stored as either a white or a black point. This type of representation is also referred to as a *binary image*.

Figure 1.1 shows a copy of a photographic print used as a standard test target for a variety of purposes in electronic document imaging systems. Figure 1.2, which shows a bit mapped image version of a portion of the same test target, was created by scanning the test target in a scanning device that produces a bit mapped representation of the original page. The scanning process will be described in detail in Chapter 2.

Figure 1.3 shows an enlargement of a section of the same bit mapped image. In the enlargement, it is possible to note that shapes and lines that appear smooth in Figure 1.2 are actually not smooth in the scanned image. Figure 1.4 shows a higher resolution enlargement of the same section, and it is possible to see the individual black and white dots that comprise the sampled digital image of the original page.

Each sampled point within the scanned image is referred to as a *picture element*, or *pixel* for short. Typically, document material is scanned so that 200 or

g drawings on aperture cards, and the long term insurance claims forms and other documents on index to film records is often created to aid in on individual physical film records. Computer support query and retrieval of specific document film roll, microfiche card, or aperture card.

Document processing systems have emerged that processing of paper or the use of micrographic processing systems perform functions similar to s. In electronic document management systems, documents is replaced by a process of electronic. Electronic document processing systems digital image format that are created by sophisticated. Once the document material is converted can be routed, displayed, stored, and retrieved computer equipment and software designed for the systems have also been used to replace micro-cases a phased transition has occurred that has ent database originally stored on film and converted format using specially designed film scanning

rocessing systems include the following types of

ment material and convert it to a digital image processing by computer compatible equipment used to store large numbers of scanned images with high resolution displays for viewing the

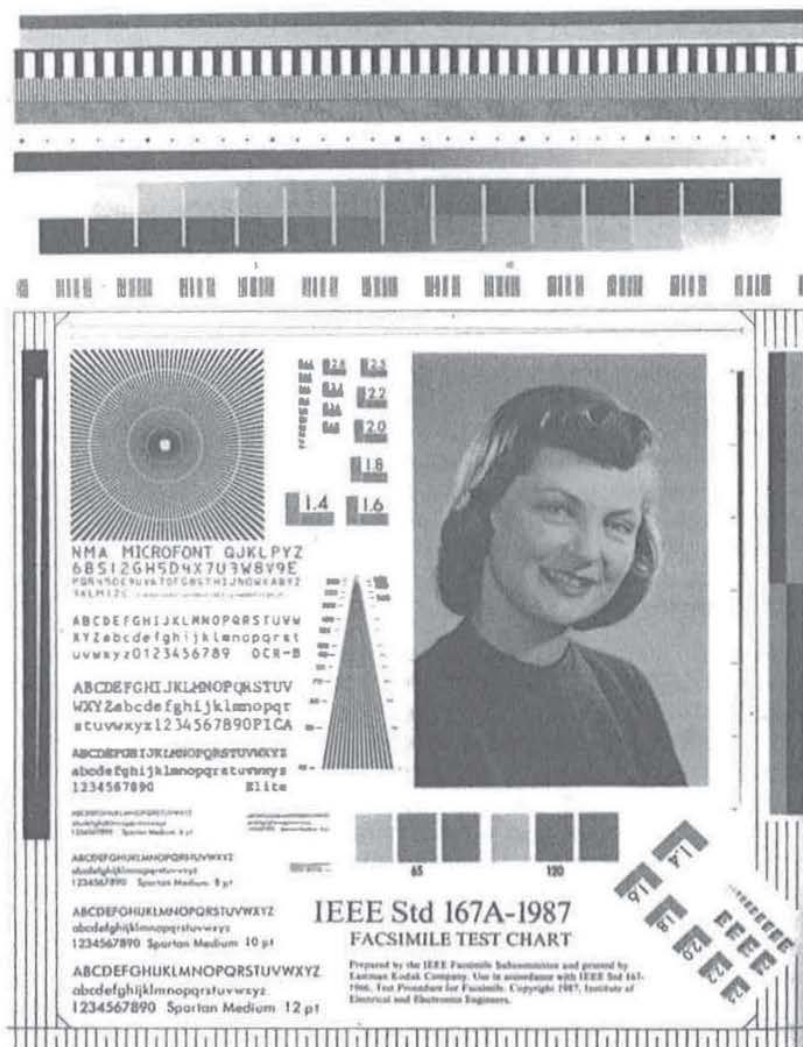
tems that transfer images between elements of

(printers) used to provide copies of the scanned

ripheral devices, including magnetic and optical netic tape subsystems

ription of the conversion of document material can be processed, manipulated, and displayed. A more complete description of the hardware by a discussion of the advantages of using an iments. A breakdown of the major categories of systems is presented, and the organization of the







### EE Std 167A-1987 FACSIMILE TEST CHART

Prepared by the IEEE Facsimile Subcommittee and printed by Eastman Kodak Company. Use in accordance with IEEE Std 167-1986, Test Procedures for Facsimile. Copyright 1987, Institute of Electrical and Electronics Engineers.

commonly used for quality control in electronic

document are acquired (although some applying resolution, depending on the resolution information in the sampled scanned image). If at 200 dots per inch, a scanned image of size

NMA MICROFONT QJKLPYZ  
6BSI2GH5D4X7U3W8V9E  
PQR45DE9UV670FG8STHIJNOWXABYZ  
3KLM12C

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
XYZabcdefghijklmnopqrst  
uvwxyz0123456789 OCR-B

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
WXYZabcdefghijklmnopqr  
stuvwxyz1234567890PICA

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxy  
z1234567890 Elite

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxy  
z1234567890 Spartan Medium 6 pt

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxy  
z1234567890 Spartan Medium 8 pt

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxy  
z1234567890 Spartan Medium 10 pt

ABCDEFGHIJKLMNPOQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxy  
z1234567890 Spartan Medium 12 pt

30x  
300  
200  
150  
125  
100  
90  
80  
70  
60  
50  
40

IE

Figure 1.2 A portion of an electronically scanned digital bit mapped version of the test target shown in Figure 1.1. The test target was scanned at a spatial resolution of 300 dots per inch. Figure provided by Cincinnati Bell Information Systems.

1,700 x 2,200 pixels is produced. Since each pixel requires one bit to represent intensity (either black or white), a total of 3.74 million bits of information is required to store a digital image representing an 8.5 x 11 inch page scanned at 200 dots per inch. If the same 8.5 x 11 inch document is scanned at 300 dots per inch, an image of size 2,550 x 3,300 pixels is produced, resulting in a total storage requirement of 8.415 million bits. Note that scanned images of smaller documents (such as credit card remittance forms, receipts, and checks) will require significantly less storage than these examples, and digitized versions of engineering drawings and larger documents will require significantly more storage for each image than in these examples. The next section of this chapter describes



NMA MICROFONT QJKLPYZ  
 6BSI2GH5D4X7U3W8V9E  
 PQR45DE9UV670FG8STHIJNOWXABYZ  
 3KLM12C

ABCDEFGHIJKLMN OPQRSTUVWXYZ  
 XYZabcdefghijklmnopqr st  
 uvwxyz0123456789 0CR-B

ABCDEFGHIJKLMN OPQRSTUVWXYZ  
 WXYZabcdefghijklmnopqr  
 stuvwxyz1234567890PICA

ABCDEFGHIJKLMN OPQRSTUVWXYZ  
 abcdefghijklmnopqrstuvwxy z  
 1234567890 Elite

ABCDEFGHIJKLMN OPQRSTUVWXYZ  
 abcdefghijklmnopqrstuvwxy z  
 234567890 Spartan Medium 6 pt

ABCDEFGHIJKLMN OPQRSTUVWXYZ  
 abcdefghijklmnopqrstuvwxy z  
 1234567890 Spartan Medium 8 pt

ABCDEFGHIJKLMN OPQRSTUVWXYZ

TF

Figure 1.3 An enlarged version of the segment shown in Figure 1.2. Note the jagged appearance of the character edges and resolution bars, showing the discrete nature of the scanned image. Figure provided by Cincinnati Bell Information Systems.

AN

pqr

par

JK

mn

Figure 1.4 and 1.3. The characters.

the comput  
 the large v  
 agement sy  
 The proc  
 resentations  
 ment proces  
 point, it is i

i) A b  
 docu  
 mate  
 map  
 the l  
 proc  
 senti





the scanning resolution is increased, because the digital scanning process is by nature a subsampling process applied to the original document material. Again, this subject is discussed in more detail in Chapter 2, and guidelines are provided regarding selection of scanning resolution to preserve the information content of the original material.

- ii) A bit mapped image will capture all information contained on the original document page. In addition to typeset material, a bit mapped image will capture other material on the original page, including signatures, handwritten information, endorsements, annotations between lines or in margins, bar codes, line drawings, lines separating sections of forms or printed material, etc. The bit mapped image representation thus provides the most complete computer compatible "replica" of the original document of any of the formats described in this section.
- iii) The bit mapped binary image format can only represent every point within the image as either black or white. The photographic insert within the test target is not well represented within the bit mapped image format. In order to represent document information that is photographic in nature, it is necessary to use digital image formats that are capable of retaining more than two intensity levels (black or white) for each sampled point within the image. The "shades of gray" present in the original material must be converted into a computer compatible format for use in electronic document management if the photographic material is important within the specific application. Electronic systems that are designed to process material such as driver's licenses, passports, and real estate records with pictures of property attached must accommodate this material. There are a variety of methods used to retain this type of information within sampled digital image formats. One technique that maintains the single bit of storage for each sampled image point is called halftoning; clusters of individual black and white dots are used to achieve a visual effect of gray shading. Halftoning preserves the original sense of gray shaded imagery, with some loss of image quality, while not appreciably increasing the amount of computer storage required for the sampled digital imagery. A second method, which preserves gray scale or color information with image quality superior to halftoning but at a cost of significantly increased computer storage per image, involves the use of more than a single bit of information to represent intensity at each sampled point within the image. These topics will be addressed in later chapters of this text. At this point, however, it is essential to note that the bit mapped binary image format is well suited to representation of document material that is commonly found in the majority of paper processing applications—document material containing textual material (printed, handwritten, or both) and/or high contrast line art. Handling of document

material in applicati  
in the original docu  
agement systems re

## Relationship to Opt

Systems are also available *character recognition (OCR)* page. These systems incor each individual printed cha characters in computer con sentation) that represent th systems also retain the gene ward that information as par

OCR systems perform w technology for handwritten recognition of printed or 1 years. Recognition rates wil size of the text, and the con machine copies, material g common type fonts by elec page material with non-whi performed on first generatio

The advantages and disac Generally, bit mapped imag important to retain comp Electronic document proces are analogous to film-base stored, retrieved, routed bet locations. Each image repre to create the digital image. from the original document- annotations, endorsements, l serves the information in th page.

OCR systems are typicall from documents. One comm entire character stream is pr tions software operating on ally a word processing or el is used to avoid a more ted operator, and a spelling corre provided by OCR systems.



increased, because the digital scanning process applied to the original document is discussed in more detail in Chapter 2, and regarding selection of scanning resolution to present of the original material.

capture all information contained on the original to typeset material, a bit mapped image on the original page, including signatures, endorsements, annotations between lines or in drawings, lines separating sections of forms or bit mapped image representation thus provides a computer compatible "replica" of the original document described in this section.

image format can only represent every point as black or white. The photographic insert within represented within the bit mapped image form document information that is photographic in use digital image formats that are capable of intensity levels (black or white) for each sample. The "shades of gray" present in the original is converted into a computer compatible format for use in document management if the photographic material is important. Electronic systems that are designed as driver's licenses, passports, and real estate property attached must accommodate this material. The methods used to retain this type of information are image formats. One technique that maintains the original image is called halftoning; black and white dots are used to achieve a visual representation of gray scale. Halftoning preserves the original sense of gray scale with a loss of image quality, while not appreciably increasing computer storage required for the sampled digital image. The method, which preserves gray scale or color with a quality superior to halftoning but at a cost of significant computer storage per image, involves the use of digital image information to represent intensity at each sample. These topics will be addressed in later chapters. However, it is essential to note that the bit mapped image is well suited to representation of document information found in the majority of paper processing applications containing textual material (printed, or high contrast line art. Handling of document

material in applications where preservation of gray shaded image material in the original document is required within the electronic document management systems requires more advanced forms of image representation.

## Relationship to Optical Character Recognition Systems

Systems are also available that can scan individual pages and utilize *optical character recognition (OCR) technology* to extract the printed material from the page. These systems incorporate pattern recognition technology that isolates each individual printed character of text and produces a series of alphanumeric characters in computer compatible format (generally in ASCII character representation) that represent the printed material on the original page. Some OCR systems also retain the general page layout or format of the original page and forward that information as part of the computer file created for each scanned page.

OCR systems perform with varying degrees of accuracy. Reliable recognition technology for handwritten characters is emerging, and commercial systems for recognition of printed or typeset characters have been available for several years. Recognition rates will vary with the quality of material being scanned, the size of the text, and the complexity of the page layout. OCR performed on copy machine copies, material generated by laser printers, material produced in less common type fonts by electronic publishing systems, multiple part forms, and page material with non-white background will achieve less accuracy than OCR performed on first generation clean printed copy with more common type fonts.

The advantages and disadvantages of the two formats are shown in Table 1.1. Generally, bit mapped image representation is used in applications where it is important to retain complete images of the original document material. Electronic document processing systems using bit mapped image representations are analogous to film-based micrographics systems. Document images are stored, retrieved, routed between work areas, and can be viewed at a variety of locations. Each image represents one side of a piece of paper that was scanned to create the digital image. The scanning process captures all the information from the original document—printed material, handwritten information, line art, annotations, endorsements, lines and spaces defining fields on forms—and preserves the information in the exact location at which it occurs on the original page.

OCR systems are typically used to capture the alphanumeric text information from documents. One common application is full page OCR capture, where the entire character stream is produced by an OCR device and forwarded to applications software operating on a host processor. The applications software is generally a word processing or electronic publishing applications package, and OCR is used to avoid a more tedious data entry process. A human quality control operator, and a spelling corrector, are often used to improve the accuracy of data provided by OCR systems.

Table 1.1 Comparison Between Binary Image and Character Representation

	Advantages	Disadvantages
Bit-Mapped Images	Replicates all information contained in original document—line art, signatures, etc.	Requires relatively large amount of computer storage for each image
Alphanumeric Representation	ASCII character data for each scanned page requires minimal amount of computer storage	Cannot capture non-textual information on document Limited accuracy on printed text; hand-written OCR still in development stage

OCR technology can also be used to capture information contained in specific fields within scanned documents. An insurance claim application might involve extraction of printed information from a series of preprinted forms. An OCR-based system can interpret an identifier that is printed in a specific area of the form to designate a particular form type. The application can then use the knowledge of the form type to extract character information from specific locations on that form, based on the form design. This process would minimize computer storage requirements, since storage of character information requires less storage than bit mapped image formats and because only the desired information is being scanned and stored in the computer system. Another example might be the use of an OCR system coupled with a full page binary image scanner, where the OCR device is used to scan and interpret a unique sequence or identification number, which is later used as an identifier to index, store, and retrieve the image of the full page in bit mapped format.

OCR technology is described briefly in Chapter 11 of this text, along with a discussion of the potential uses of OCR technology within a system primarily used for management of document information in bit mapped image format. The main focus of this text, however, will be on electronic document management systems that are based on utilization of bit mapped image format for image representation of the source documents within an application. The use of full image representation of scanned document material is a more flexible technology than OCR because of the wide range of information that is preserved using the bit mapped image format and the preservation of the exact shape and location of the information content of the original source document. Full page bit mapped imagery is the electronic equivalent of a photographic copy of a document used in micrographic systems.

OCR systems can be highly effective in applications whose requirements can be met with the limited information representation capabilities of OCR technology. OCR technology is generally used in applications for which the following constraints are acceptable:

- i) inform  
nology  
alphan
- ii) Minim
- iii) No req  
of scan
- iv) No req  
endorse
- v) No req  
extracte

## ELECTRO

### Enabling Te

The developme  
been made poss  
nologies. These

#### Economical

page of a docum  
bits of storage f  
of computer tecl  
of the write on  
was a key techn  
management sy  
devices that mir  
made it possible  
document materi  
media that have  
developed very t  
storing thousand  
within reasonab  
speeds for retriev

#### Electronic Sc

first began to b  
applications in  
become comm  
systems. This te  
source documen  
power scanning  
component is co



## Binary Image and Character Representation

Advantages	Disadvantages
Replicates all information contained in original document—line art, signatures, etc.	Requires relatively large amount of computer storage for each image
ASCII character data for each scanned page requires minimal amount of computer storage	Cannot capture non-textual information on document Limited accuracy on printed text; hand-written OCR still in development stage

used to capture information contained in specific s. An insurance claim application might involve m from a series of preprinted forms. An OCR-identifier that is printed in a specific area of the form type. The application can then use the xtract character information from specific loca-form design. This process would minimize come storage of character information requires less orms and because only the desired information ie computer system. Another example might be ed with a full page binary image scanner, where nd interpret a unique sequence or identification an identifier to index, store, and retrieve the ed format.

briefly in Chapter 11 of this text, along with a of OCR technology within a system primarily it information in bit mapped image format. The , will be on electronic document management on of bit mapped image format for image rep-its within an application. The use of full image nt material is a more flexible technology than of information that is preserved using the bit ervation of the exact shape and location of the al source document. Full page bit mapped at of a photographic copy of a document used

ective in applications whose requirements can n representation capabilities of OCR technol-used in applications for which the following

- i) information content can be recognized with high accuracy by OCR technology, minimizing the requirements for quality control on the scanned alphanumeric information
- ii) Minimal amount of hand printed information
- iii) No requirement for retaining line drawings and/or photographic content of scanned documents
- iv) No requirement for capture of information contained in written notations, endorsements or stamps entered on the original source document material
- v) No requirement to preserve the geographic location of information extracted from the page being scanned

## ELECTRONIC DOCUMENT PROCESSING SYSTEMS

## Enabling Technology

The development of cost effective electronic document processing systems has been made possible through the evolution of several key computer science technologies. These technologies include the following:

**Economical Compact Digital Storage Technology.** Representation of a page of a document as a scanned digital bit mapped image requires thousands of bits of storage for each page of scanned material. Although limited applications of computer technology to document management existed prior to the evolution of the write once read many (WORM) optical disk storage technology, WORM was a key technology in establishing the viability of cost effective electronic management systems. WORM technology, and other compact digital storage devices that minimize the cost per bit of storage for large quantities of data, made it possible to develop systems for managing realistically large volumes of document material on computer based systems. WORM disk, and other compact media that have emerged within the computer industry (including recently developed very high density digital tape storage systems) have the capacity for storing thousands of pages of information at a reasonable cost per bit of storage, within reasonable areas of floor space, and they provide reasonable access speeds for retrieving document material.

**Electronic Scanning Technology.** Solid state electronic image scanners first began to be utilized in the space program and in earth observations applications in the 1980s. More recently, solid state imaging systems have become common in commercial television cameras and in photographic systems. This technology is also used in the electronic scanners that convert source document material to bit mapped image representation. A low cost low power scanning array located within a high speed solid state electronic component is coupled with an illumination source, a mechanical transport, and



an optical system to create a modern electronic scanning system. Compact, minimal cost, high performance electronic scanning systems enable development of systems that can capture large daily volumes of arriving paper material and/or efficiently convert a stored paper archive into digital format in a reasonable amount of time.

**High Speed Communications Technology.** The evolution of modern computing systems from highly centralized systems designed with a single processor and a limited number of simultaneous users to modern distributed systems involving networked connection of hundreds or thousands of individual computing nodes has driven the development of high speed communications capability within the computer industry. Commercial systems have been developed to transfer large volumes of data at high speed within modern computing systems. These systems incorporate a series of hardware components specially designed to transfer data at high speed. A series of industry standards has evolved governing interchange of data between systems, which may be procured from multiple vendors. The standards govern both the hardware and software in commercial systems. A set of established protocols is used to aggregate data into standardized packets, and they govern highly reliable data transfer at relatively high speeds between multiple locations within distributed computer systems. The volume of data associated with digitized imagery is relatively large, compared to the volume of data involved in most commercial applications that are generally more text oriented than image oriented. The communications link is often the most constraining data transfer link within an electronic document management system. Despite that fact, commercial communications equipment and networking standards are an important enabling technology for electronic document management systems. In recent years, vendors have begun to realize that increased image traffic within distributed computer environments will be a fact of life in the future. Current efforts are focused on design and definition of standards governing high speed efficient transfer of image data within networked systems.

**Image Display Technology.** Effective display of digitized document imagery requires high resolution display technology. Although small forms, such as credit card remittance slips and checks, can be displayed on systems supporting display of a small number of picture elements, display of images obtained by scanning 8.5 x 11 inch paper requires image display equipment capable of displaying at least 1,700 x 2,200 pixels if the page is scanned at 200 dots per inch. The commercial evolution of high resolution display systems to support other requirements has provided the technology required for display of full page scanned document material for electronic document management systems. Hardware that supports "roam" and "zoom" functions has evolved to provide the ability to display the full image at reduced resolution and then magnify selected portions of the document material to full resolution for detailed work. This technology supports processing of large format drawings using

electronic systems. High resolution scanning systems enable development of systems that can capture large daily volumes of arriving paper material and/or efficiently convert a stored paper archive into digital format in a reasonable amount of time.

**Image Compression.** Image compression, as described, provides a method to accommodate large volumes of data transfer time. Document management systems have been developed for fast data transfer time. Compression is optimal for any facsimile world; and decompression resulted in wide document processing. Compression digitizes material in the most detailed information better than image minimal "white space" discussed in detail.

## System Hardware

The components shown in Figure 1

**Image Scanning.** Image scanning material into scanned images. Available scanner media. Paper scanning documents. There are (checks, credit cards, office size documents to 11 x 17 inch paper drawings). Commercial paper handling methods are addressed in detail.

Another class of scanners are designed for high speed scanning. They have been converted to high speed (typically used for



a modern electronic scanning system. Compact, electronic scanning systems enable development of the daily volumes of arriving paper material and/or paper archive into digital format in a reasonable

**Communications Technology.** The evolution of modern highly centralized systems designed with a single user of simultaneous users to modern distributed connection of hundreds or thousands of individual users. The development of high speed communications in the industry. Commercial systems have been able to handle volumes of data at high speed within modern systems incorporate a series of hardware components that transfer data at high speed. A series of industry standards govern the range of data between systems, which may be different. The standards govern both the hardware and software. A set of established protocols is used to transfer data packets, and they govern highly reliable data transfer between multiple locations within distributed systems. The volume of data associated with digitized imagery is much more than image oriented. The volume of data involved in most commercial systems is more text oriented than image oriented. The most constraining data transfer link within an electronic document management system. Despite that fact, commercial networking standards are an important enabling technology for document management systems. In recent years, that increased image traffic within distributed systems is a fact of life in the future. Current efforts are in the area of standards governing high speed efficient electronic document management systems.

Effective display of digitized document material is a display technology. Although small forms, checks, and slips, can be displayed on systems that transfer a number of picture elements, display of images of a full page requires image display equipment that transfer 30 x 2,200 pixels if the page is scanned at 200 lines per inch. The evolution of high resolution display systems to provide the technology required for display of material for electronic document management systems. "Roam" and "zoom" functions has evolved to allow a full image at reduced resolution and then document material to full resolution for detailed processing of large format drawings using

electronic systems, and it also supports display of images obtained by high resolution scanning of standard office size documents, for applications where the high resolution scanning is required to retain the information in the original document.

**Image Compression.** Digital images require significant amounts of computer storage, as described earlier in this chapter. Image compression technology provides a method for reducing the amount of storage capacity required to accommodate scanned images. Transfer of compressed images also reduces the data transfer time across communications systems and networks. Most electronic document management systems utilize a set of compression standards originally developed for facsimile systems, driven by the need to minimize telephone line transfer time between "fax" equipment. These standards are not necessarily optimal for any specific application, but the fact that standards did exist in the facsimile world and special purpose hardware could be developed to compress and decompress digitized document images based on these standards, has resulted in widespread use of the facsimile compression standards in the document processing world. For binary images, these standards can typically compress digitized image by factors of from 2 to 10, depending on the type of material in the scanned image. (Material with lower amounts of extremely detailed information, such as typed pages of text mixed with line art compress better than images of pages containing large amounts of extremely small type, minimal "white space", and no line art, for example). Image compression is discussed in detail in Chapter 10.

## System Hardware Components

The components that make up an electronic document management system are shown in Figure 1.5. Systems usually include the following components:

**Image Scanners.** These are specialized devices that convert document material into sampled digital imagery. There are a variety of commercially available scanners designed to scan document material from paper and film media. Paper scanners are designed to handle various sizes and weights of paper documents. There are scanners designed to capture images of small forms (checks, credit card slips, adding machine tapes), scanners designed to handle office size documents (ranging from 3 x 5 inch cards and 8.5 x 11 inch pages up to 11 x 17 inch pages), and scanners designed to handle large engineering drawings. Commercially available scanners vary greatly in their scanning speed, paper handling mechanisms, and the output image quality that is produced. This topic is addressed in more detail in Chapter 2.

Another class of commercially available scanner is the film scanner. Film scanners are designed to produce sampled digital imagery of documents that have been converted into filmed images. Film formats can include aperture cards (typically used for engineering drawings, these are computer compatible cards



with film inserts containing document images), 16 mm microfilm in various formats, and 105 mm microfiche cards that contain multiple document images in a two dimensional array.

A single electronic document management system may contain several types of scanners. It might have a high throughput centralized scanner for production scanning, several lower throughput desktop scanners distributed across several work areas to capture a low volume of document material each day, and one or more film scanners that might be used to scan an existing micrographics database that has never been completely converted to a digital data base.

**High-speed Communications Interface.** Most electronic document management systems support multiple users in a distributed environment. The communications components in an electronic document management system must connect multiple users, and a variety of devices and subsystems (image data storage equipment, scanners, printers, workstations, external communications interfaces, etc.). At any given time, there are generally several overlapping requests pending in the system for transfer of image data. Several users may be requesting several different images for use at their individual workstations, a print request from one user may be pending (routing an image from a display workstation to the laser printer), and the production scanner may be sending scanned images to the data base server for indexing and storage. The data

transfer rates require demands generally a pathway. In many (such as Ethernet<sup>®</sup> m associated with trans system within electri the design effort, an lead to poor respons

**Mass Storage.** I document manager magnetic tape. Magi writing information, options include WO Read Only Memor systems usually incl used for temporary s the location in dig applications softw document data bases can be used for both High density magnet backup media for o facilities. Mass stor management systems centralized mass stor.

**Displays.** Image the sampled binary is effective in displayin only alphanumeric i imagery, the resolutio may not be sufficient an 8.5 x 11 inch page pages scanned at 20 available display : management system display equipment fo data may be stored a compressed format to decompression hardw computer workstation

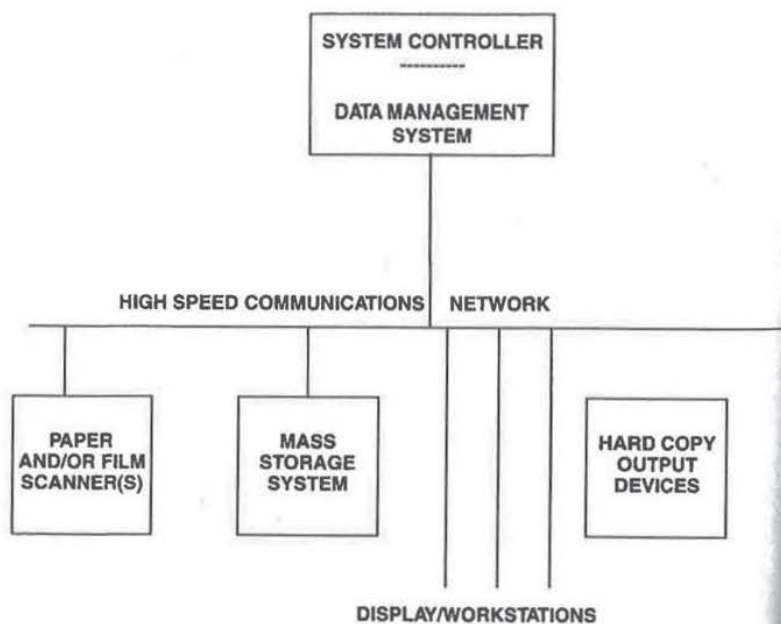


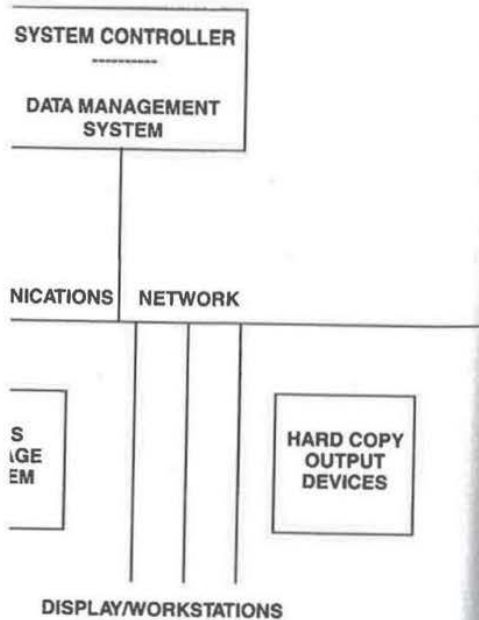
Figure 1.5 Typical components of an electronic document management system.

\* A registered trademark

cument images), 16 mm microfilm in various formats that contain multiple document images in a

an electronic document management system may contain several types of high throughput centralized scanner for production throughput desktop scanners distributed across several volumes of document material each day, and one or more may be used to scan an existing micrographics database and convert it to a digital data base.

**Communications Interface.** Most electronic document management systems support multiple users in a distributed environment. The system is composed of a variety of devices and subsystems (image data base, printers, workstations, external communications interface). At any one time, there are generally several overlapping communications paths for transfer of image data. Several users may be viewing images for use at their individual workstations, a user may be pending (routing an image from a display to a printer), and the production scanner may be sending image data to a base server for indexing and storage. The data



an electronic document management system.

transfer rates required to support an acceptable system response time to multiple demands generally dictate the use of a high speed distributed communications pathway. In many document management applications, commercial networks such as Ethernet\* may not be adequate to support the communications demands associated with transfer of document image data. Design of the communications system within electronic document management systems is a very critical part of the design effort, and a design based on a low estimate of the image traffic can lead to poor response times and unproductive users.

**Mass Storage.** Mass storage components in common use within electronic document management systems include magnetic disk, optical disk, and magnetic tape. Magnetic disk provides the fastest access time for reading and writing information, but is the lowest density of the three media. Optical disk options include WORM (Write Once Read Many), CD-ROM (Compact Disk Read Only Memory), and erasable optical disks. Document management systems usually include all three types of media. Typically, magnetic disk is used for temporary storage, such as the index information or catalog containing the location in digital storage of every document, operating system and applications software. WORM is used for long term archival storage of document data bases that do not change dynamically, and erasable optical drives can be used for both archival and transient image storage at very high density. High density magnetic tape can be used as an alternative to optical storage, as a backup media for optical disks, and as a physical exchange media between facilities. Mass storage can be centralized or distributed within document management systems, and it is possible that the same image can reside on both a centralized mass storage system and on local storage at a workstation.

**Displays.** Image display devices are designed to present a visual display of the sampled binary imagery to the user. Standard computer displays may not be effective in displaying document information. They may be designed to display only alphanumeric information. If they are designed to display bit mapped imagery, the resolution (the number of picture elements displayed on the screen) may not be sufficient to display a full digital image. Sampled digital imagery of an 8.5 x 11 inch page may require screen resolutions of 1,700 x 2,200 pixels for pages scanned at 200 dpi, as discussed previously, and many commercially available display systems may not provide this capability. Document management systems typically include specially designed high resolution display equipment for display of document material. In addition, since the image data may be stored and transferred across the network to the display system in compressed format to minimize transfer time. The display system may include decompression hardware. Image display systems thus tend to be software driven computer workstations, sometimes augmented with special purpose hardware.

\* A registered trademark of Xerox Corporation.



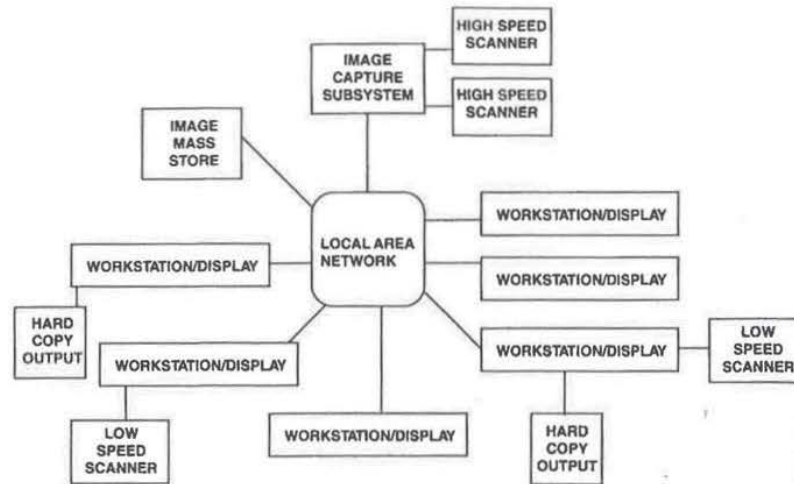


Figure 1.6 A typical electronic document management system.

**Hard Copy Output Devices.** It is often necessary to generate paper hard copy of the document material managed by an electronic system. The most common hard copy output device is laser printers because they provide sufficient resolution for reproduction of scanned binary imagery. Typical laser printers provide 300 dots per inch output capability, which is often a good match for digitized document material that has been scanned at or near that resolution.

Figure 1.6 shows a detailed diagram of a typical electronic document management system. The figure shows that there are many distributed workstations linked by a common network. Some functions are supported centrally (such as the image mass storage) while other functions are supported in a distributed fashion (such as hard copy output). The scanning function in the system shown in Figure 1.6 is also distributed, with a high speed production image capture system located centrally, and lower capacity scanners distributed around other points in the system.

### System Software Components

An electronic document management system will include several levels of software, as shown in Figure 1.7. The typical software components will include the following:

**Operating System.** The basic computer operating system provides overall control of processing and input/output functions. It may also provide the user

interface with the document management system. The system may include DOS (Data Operating System) or other commercially available operating systems. The system may also provide standard electronic document management functions. General system and providing

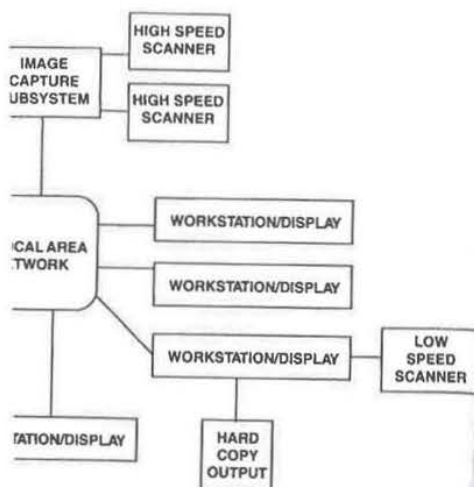
**Applications Software.** Specific document management systems are provided to support insurance claim applications code facility that could software generally and retrieval, image types will be provided. The system may be situated in the system. For by a workstation and indexing that



Figure 1.7 Software system.

- \* A trademark of N
- \*\* A trademark of I
- \*\*\* A trademark of A





Document management system.

It is often necessary to generate paper hard copy of an electronic system. The most common method is to use laser printers because they provide sufficient resolution to reproduce scanned binary imagery. Typical laser printers have a resolution of 300 dots per inch, which is often a good match for the resolution of the original document as been scanned at or near that resolution.

A diagram of a typical electronic document management system shows that there are many distributed workstations where some functions are supported centrally (such as scanning) and other functions are supported in a distributed manner. The scanning function in the system shown is supported by a high speed production image capture system with capacity scanners distributed around other

## Components

A document management system will include several levels of software. Typical software components will include the

computer operating system provides overall control and input functions. It may also provide the user

interface with the system as well, depending on the design of the electronic document management system. The operating systems most commonly used include DOS (Disk Operating System)\*, OS/2,\*\* and Unix.\*\*\* These are all commercially available operating systems supported by multiple vendors that provide standardized environments for development of applications software. In most electronic document management systems, the end user of the system may not know the specific operating system and may never have to interact with the basic computer operating system in order to perform document management functions. Generally, the applications software is built on top of the operating system and provides the user interface support.

**Applications Software.** The applications software is designed to support the specific document processing requirements of the particular application. Systems are provided with either general purpose applications software that is developed to support a specific application market (such as check processing, insurance claim processing, or archival storage and retrieval) or with applications code that has been custom designed for a specific application or facility that could not be easily exported to another organization. Applications software generally supports document acquisition and indexing, image storage and retrieval, image display, and hard copy generation. User interfaces of various types will be provided for each of these operations. Selected components of the system may be supported with specific software that is not available elsewhere in the system. For example the centralized scanning function may be controlled by a workstation with specific software designed to support document scanning and indexing that is not available on other workstations in the system.

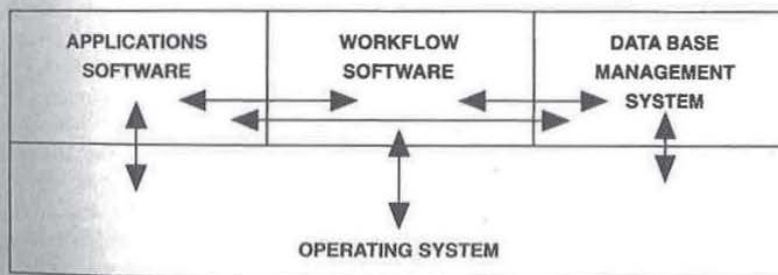


Figure 1.7 Software structure within a typical electronic document management system.

\* A trademark of Microsoft Corporation

\*\* A trademark of IBM Corporation

\*\*\* A trademark of AT&T, Inc.

**Workflow Software.** In the past few years, the concept of workflow software has become important in electronic document management systems. Many applications involve a flow of document processing similar to that shown in Figure 1.8. Documents are scanned on a periodic basis (hourly, daily, overnight, etc.), indexed and stored on the image mass store. From that point, sets of documents may be aggregated and forwarded to individual workstations for review and processing. Processing may proceed in several steps, involving routing of the document from one area to another (from one workstation to another). A claim form may be reviewed at one workstation for example, and the proposed resolution of the claim may be reviewed by a supervisor at a second workstation. A third workstation may be involved in generating correspondence or issuing the payment associated with the claim. The software that controls the flow of documents through this type of system is called workflow software. Workflow software can also include software that provides status information for management, as well as statistics on throughput and operator performance.

**Data Base Management Software.** This class of software is often found integrated within the applications support software in electronic document management systems. In most systems, there is an indexing scheme designed for the particular application. In one system, document material may be structured logically into *folders*, *documents* within folders, and *pages* within documents. In this example, there are three levels of indexing. Documents are comprised of one or more pages, and folders are comprised of one or more documents. In another system, each page of material that is scanned might be assigned a unique identifier or index; a check processing application is an example where only a single level of indexing may be required. The data base management system (DBMS) makes it possible to query the document data base and retrieve information based on the logical structure of the index. The DBMS makes it possible to retrieve specific documents or folders in a three level indexing scheme or individual images in a single indexing level application. The DBMS can generally accommodate more than a single index at each level. For example, documents may be indexed based on document type, date of origin, and other parameters. The DBMS also makes it possible to retrieve documents based on multiple search criteria—for example locate all documents of form type B generated after a specific date.

The categories listed above are typical categories of software commonly found in document management systems. Specific systems may incorporate additional software categories, depending on the application, but these are the categories most commonly found in today's systems.

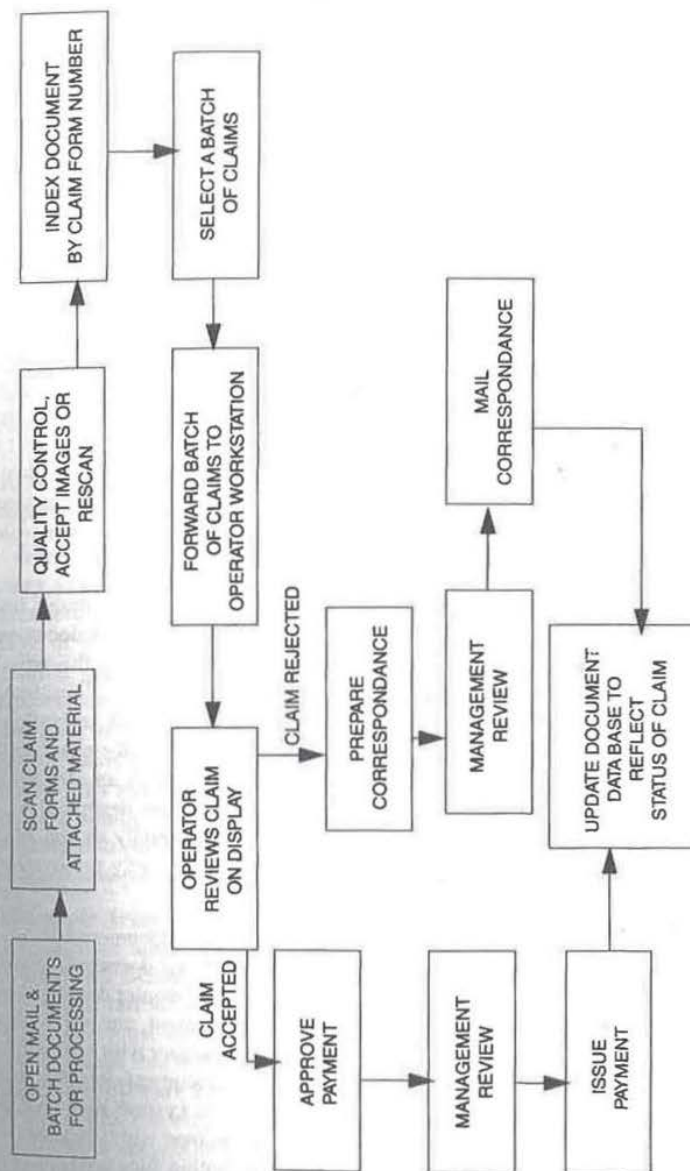




the past few years, the concept of workflow software electronic document management systems. Many of document processing similar to that shown in scanned on a periodic basis (hourly, daily, overnight, on the image mass store. From that point, sets of data are created and forwarded to individual workstations for processing may proceed in several steps, involving moving from one area to another (from one workstation to another) to be reviewed at one workstation for example, and then the claim may be reviewed by a supervisor at a third workstation may be involved in generating the payment associated with the claim. The software used to process documents through this type of system is called workflow software can also include software that provides management, as well as statistics on throughput and

**Software.** This class of software is often found in systems that support software in electronic document systems, there is an indexing scheme designed for one system, document material may be structured into *units* within folders, and *pages* within documents. In levels of indexing. Documents are comprised of folders are comprised of one or more documents. In material that is scanned might be assigned a unique processing application is an example where only a be required. The data base management system to query the document data base and retrieve ical structure of the index. The DBMS makes it documents or folders in a three level indexing n a single indexing level application. The DBMS ore than a single index at each level. For example, sed on document type, date of origin, and other makes it possible to retrieve documents based on example locate all documents of form type B

are typical categories of software commonly found in modern systems. Specific systems may incorporate some or all of these, depending on the application, but these are the building blocks of today's systems.



**Figure 1.8** Document flow in a typical claims processing application.



## ADVANTAGES OF ELECTRONIC DOCUMENT PROCESSING SYSTEMS

There are many reasons that computer based technology is an attractive technology for managing document materials. This section enumerates some of those reasons.

**Security.** Electronic document management systems are generally a more secure way to store and process document information than systems based on paper handling or micrographics technology. The use of computer based technology means that the document material is stored digitally on computer compatible media in locations where personnel access is often limited. The media (magnetic or optical disk) is generally located within components of the system where access is difficult or limited, and physical removal of the document material might be quite difficult. The applications software can be used to limit computer based access to document material; specific individuals can be granted access to limited portions of the data base if required, and system access can be controlled by standard computer access techniques (e.g., password protection, multiple protection levels for file structures, etc.).

**Data Base Integrity.** One common problem with paper file systems, and with some film-based micrographics systems, is improper refileing of document material that has been removed. The document can be refiled in the wrong location, which makes it extremely difficult to re-locate the document. Electronic document management systems utilize modern DBMS technology and effectively eliminate this problem. Although it is theoretically possible to destroy digital data files, it is much more difficult to do so in an electronic system. In computer based systems, even if the active files are destroyed by a rare occurrence or by malicious users, a digital backup generally exists. With proper management procedures, it is very difficult to permanently remove and destroy files from a computer based system.

**Document Integrity.** Once a digital image version of a document has been created, it is extremely difficult to modify the content of that document image. Modification of electronically stored digital imagery would require development of a sophisticated set of software designed for malicious intent, and would also require direct access to the digital system by an authorized user with appropriate access privileges. It is much easier to modify a written document on paper than to modify a scanned digital image, and the electronic system provides an increased level of data integrity. The order of pages contained within documents consisting of multiple pages, and the document order within files consisting of multiple documents is preserved automatically in electronic document management systems following the scanning process that creates the electronic imagery. If the document imagery is scanned in the proper order, the order is

automatically preserve from documents, and paper, microfiche can misfiled or become file system, but electronic

**Archival Storage.** term data archives. Of with projected lifetime systems and magnetic Computer based system redundant storage of b the information is acc technology. Digital fil Digital files can also b the end of projected li the media is beginning more reliable long term

**Image Quality.** E many years of devel Commercially availab range of processing c material generated fr improves on the origin generated, electronic i cups of coffee); the im created. Any physical c standard computer bac guaranteed to maintain susceptible to permaner

**Multiple Access.** I based system and rou several users located at their workstations si documents must be p approval steps that do workflow software r workstations for sim processing a document to do this in a system individual film recor locations can be phy technology.



## 5 OF ELECTRONIC DOCUMENT PROCESSING SYSTEMS

at computer based technology is an attractive technology. This section enumerates some of those

document management systems are generally a more process document information than systems based on graphics technology. The use of computer based document material is stored digitally on computer systems where personnel access is often limited. The disk is generally located within components of the difficult or limited, and physical removal of the e quite difficult. The applications software can be d access to document material; specific individuals ited portions of the data base if required, and system standard computer access techniques (e.g., password n levels for file structures, etc.).

ne common problem with paper file systems, and graphics systems, is improper refiling of documents. The document can be refilled in the wrong extremely difficult to re-locate the document. eament systems utilize modern DBMS technology ; problem. Although it is theoretically possible to is much more difficult to do so in an electronic /systems, even if the active files are destroyed by a ous users, a digital backup generally exists. With es, it is very difficult to permanently remove and based system.

e a digital image version of a document has been It to modify the content of that document image. stored digital imagery would require development are designed for malicious intent, and would also ital system by an authorized user with appropriate isier to modify a written document on paper than image, and the electronic system provides an . The order of pages contained within documents nd the document order within files consisting of rved automatically in electronic document the scanning process that creates the electronic ery is scanned in the proper order, the order is

automatically preserved by the electronic system. Individual pages are not lost from documents, and documents do not disappear from file folders. Pieces of paper, microfiche cards, rolls of microfilm, and aperture cards can all be misfiled or become filed out of order after continuing use of a document filing system, but electronic image files automatically retain their logical structure.

**Archival Storage.** Digital data can be managed to create extremely long term data archives. Optical disk media are designed for long lifetimes (media with projected lifetimes of 30 years or more are common). Magnetic disk systems and magnetic tape systems are also designed for long term data integrity. Computer based systems also incorporate error correcting technology, based on redundant storage of binary information and continual error checking each time the information is accessed. Digitized document images benefit from use of this technology. Digital files can be protected through standard backup procedures. Digital files can also be copied to new media when existing media are reaching the end of projected lifetime and/or the error correction processes indicate that the media is beginning to degrade. For these reasons, digital storage provides more reliable long term archival storage potential than either paper or film media.

**Image Quality.** Electronic document management systems benefit from many years of development of digital image enhancement technology. Commercially available paper and film scanning systems incorporate a wide range of processing capability designed to optimize the quality of the image material generated from the original source documents. Image quality that improves on the original material is often generated by electronic scanners. Once generated, electronic images do not fade or darken (or even experience spilled cups of coffee); the image quality remains as good as on the day the image was created. Any physical damage to a digitized image data base can be overcome by standard computer backup and operational procedures. Paper and film are not guaranteed to maintain image quality over long periods of time, and they are more susceptible to permanent physical damage than digital image data records.

**Multiple Access.** Digital data files can be stored centrally within a computer based system and routed to multiple locations as required. It is possible for several users located at different physical locations each to view a document on their workstations simultaneously. This feature is especially useful when documents must be processed through a series of review, processing, and approval steps that do not have to be performed sequentially. In those cases, the workflow software may route copies of the digitized imagery to multiple workstations for simultaneous processing, and it can track the status of processing a document at multiple locations within the system. It is impossible to do this in a system that involves processing individual pieces of paper or individual film records of document material. Note that the digital work locations can be physically distant, linked by modern communications technology.



**Routing Speed.** Networking and communications technology make it possible to route document material at high speed both locally or over long distances. Document processing that involves multiple sequential steps performed at different work locations can proceed based on high speed electronic transfer of document images. There is no need to wait for cartons of interoffice mail to be transferred between floors, buildings, or facilities of an organization when using an electronic system. Using electronic technology, document images can be routed to the next work location in minutes instead of hours or days.

**Query and Retrieval Speed.** Many document management systems feature customer inquiry support functions. Customers may call and request information or status of their account. Use of electronic document management technology, including DBMS technology, can minimize the response time to these queries. It is generally much faster to locate a specific file on a computer based system than it is to find a document in a series of filing cabinets or on one of many rolls of microfilm. An operator can query and retrieve a document for display at a workstation while fielding a customer inquiry, providing an advanced level of customer support that is difficult to achieve using non-electronic means.

## ORGANIZATION OF THE TEXT

This text is organized into two sections. The first section (Chapters 1 through 7) provides a detailed introduction to the components used in electronic document management systems with a minimal amount of detailed technical information. Readers interested in an overview of the major system components can read these chapters and acquire adequate knowledge to understand basic concepts, the type and variety of general purpose and specialized hardware used in these systems, and the general decisions that must be made when designing or purchasing an electronic document management system. Chapter 2 describes the process of image capture, including a description of the most common types of paper and film scanners and scanning issues that determine image quality retained in the digitized imagery. Chapter 3 provides an introduction into the systems used to display digitized images and the functions that can be incorporated within image displays. Chapter 4 describes board level components that are widely used in electronic document management systems. These products are designed to complement personal computer or workstation systems and provide high performance specialized document management functions (data compression and decompression, high resolution display control, etc.). Chapter 5 describes optical disk storage technology, including a summary of the various types of optical disk media and the devices used to store, retrieve, read, and write optical media. Chapter 6 provides an outline of system design issues that must be resolved in designing or purchasing a system for a specific application. Chapter 7 describes a variety of actual systems in operation today.

working and communications technology make it possible to move material at high speed both locally or over long distances. Processing that involves multiple sequential steps at different work locations can proceed based on high speed transmission of images. There is no need to wait for cartons of material to be moved between floors, buildings, or facilities of an electronic system. Using electronic technology, documents can be moved to the next work location in minutes instead of hours.

4. Many document management systems feature remote access. Customers may call and request information. The use of electronic document management technology can minimize the response time to these queries. It is faster to locate a specific file on a computer based system than to search through series of filing cabinets or on one of many rolls of microfilm. A query and retrieve a document for display at a workstation. A customer inquiry, providing an advanced level of service, can be achieved using non-electronic means.

## ORGANIZATION OF THE TEXT

The text is organized into sections. The first section (Chapters 1 through 7) is introductory, covering the components used in electronic document management systems. Chapter 1 provides a minimal amount of detailed technical information. Chapter 2 provides a review of the major system components that can be read. Chapter 3 provides adequate knowledge to understand basic concepts, such as the purpose and specialized hardware used in these systems. Chapter 4 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 5 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 6 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 7 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 8 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 9 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 10 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 11 describes the design and construction of systems that must be made when designing or purchasing a document management system. Chapter 12 describes the design and construction of systems that must be made when designing or purchasing a document management system.

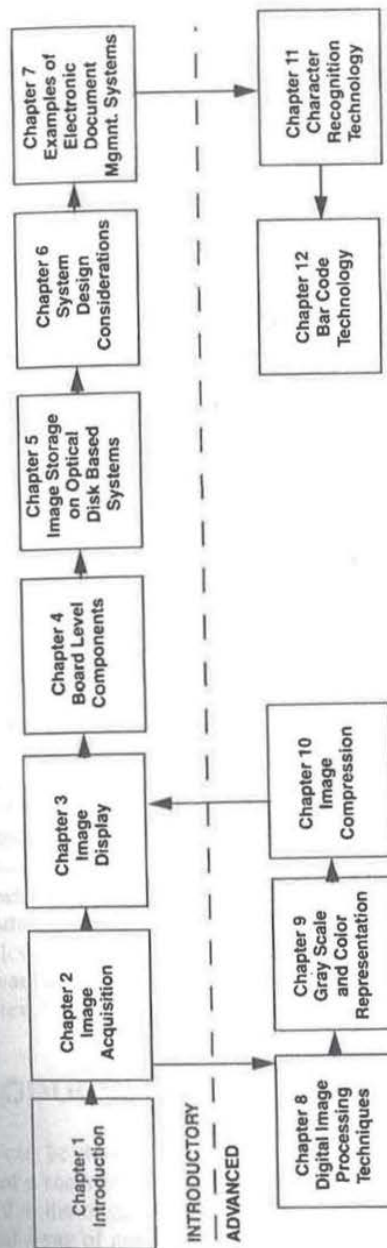


Figure 1.9 Organization of the text.



Chapters 8 through 12 deal with more specialized topics in more depth. The reader interested in additional detail on the topics covered by the advanced topics section of the text can read individual chapters of interest after completing the first seven chapters. An alternate approach is shown in Figure 1.9. Readers not interested in the more detailed information can read Chapters 1 through 7 in sequence; those interested in technical detail may follow an alternate path, reading Chapters 8, 9 and 10 between Chapters 2 and 3, and then reading Chapters 11 and 12 after completing Chapter 7.

References to more detailed information are also contained in selected chapters. The Appendix contains a general list of standard references, technical conferences, journals, and other sources of information available to those interested in keeping up to date in a rapidly changing area of new technology.

This chapter  
tal imagery  
The concep  
that determ  
discussion o

This chap  
ated with i  
into the eff  
image quali  
end of the  
more techni  
gray scale a  
used to redu  
presented in  
the advance  
to read the  
remainder o

IN

A digital im  
simple exam  
white backg  
a two dimen



## CHAPTER

## 6

# System Design Considerations

The previous chapters have provided an introduction to electronic document management systems and a detailed overview of the major components found in these systems. This chapter describes some of the system design considerations involved in developing and operating a system, and Chapter 7 describes several operational systems that address various applications.

Electronic document management systems are implemented to automate specific functions within a business, financial institution, or government agency. There are broad classes of system types designed to address specific types of applications. Systems in operation today range from those with hardware selected and software designed and developed specifically for an application to systems acquired from commercial vendors or system integrators that are applied to specific applications with no modification of basic hardware architecture or applications software. Many systems fall in between these two extremes, consisting of a commercially available system that has been adapted or customized to meet a specific application.

It is not possible to review all the detailed design and cost tradeoffs involved in developing an electronic document management system in a single chapter of a textbook. Applications differ widely and technology is evolving rapidly; these two factors make it difficult to present a hard and fast generalized set of design rules. This chapter presents an introduction to considerations involved in designing an electronic document management system. The development of a set of system objectives is described. The cost benefit analysis used in making a decision to utilize electronic document management technology to support all or part of an existing operation is reviewed. Following that review, several different generic applications are studied to determine the design and cost drivers inherent



in the applications. The chapter concludes with a review of some of the key decisions that are made while either designing or selecting a commercially available system for use in a particular application. By reviewing the topics and detailed examples presented in this chapter, and the design of the operational systems described in Chapter 7, the reader should acquire sufficient understanding to proceed with the first steps involved in (a) determining if an electronic document management system could benefit an existing operation and (b) the development of the approximate cost of implementing and operating such a system.

### **DETERMINING SYSTEM OBJECTIVES AND PERFORMING THE COST BENEFIT ANALYSIS**

The first step involved in making the decision on whether or not to implement an electronic document management system often involves a detailed examination of current operational costs for a particular process and an analysis of the cost benefit to be derived from transition to an electronic document management system. The analysis compares the cost of current operations plus future anticipated growth in current operations against the amortized cost of procurement of a computer based system, cost of training, and new operations costs based on the new system. The consideration is made over a fixed period of time in the future, and the analysis is often referred to as a *cost benefit analysis*. The cost benefit analysis may not be the only factor in making the decision to proceed with system implementation, but it is almost always a major item of consideration in the decision. It may be desirable, for example, to implement an electronic document management system to improve customer response time or data security for a particular application, even if there are additional costs associated with the implementation.

The cost benefit analysis generally starts with a complete definition of the operations and procedures within a particular activity. A review of the flow of documents within an activity, and the actions taken at every step of the processing by staff personnel, is made. This analysis results in an understanding of the *workflow* within a particular application. The review of established processes and procedures will generally result in the identification of one or more areas that are candidates for automation based on electronic document management technology. Once the workflow is established for a particular activity, it is possible to proceed to a top-level design of an electronic document management system that can be used to support that activity, or to evaluate commercially available systems against the processing requirements that have been identified.

Chapter 1 summarized some of the advantages gained by transition to electronic document management technology. Systems are generally implemented

with a review of some of the key  
ing or selecting a commercially avail-  
ation. By reviewing the topics and  
r, and the design of the operational  
ould acquire sufficient understanding  
a) determining if an electronic docu-  
xisting operation and (b) the develop-  
g and operating such a system.

## OBJECTIVES AND BENEFIT ANALYSIS

sion on whether or not to implement  
n often involves a detailed examina-  
ular process and an analysis of the  
an electronic document management  
current operations plus future anti-  
the amortized cost of procurement of  
nd new operations costs based on the  
r a fixed period of time in the future,  
ost benefit analysis. The cost benefit  
ing the decision to proceed with sys-  
s a major item of consideration in the  
to implement an electronic document  
response time or data security for a  
additional costs associated with the

ts with a complete definition of the  
lar activity. A review of the flow of  
ns taken at every step of the process-  
sis results in an understanding of the  
The review of established processes  
identification of one or more areas  
on electronic document management  
ed for a particular activity, it is possi-  
lectronic document management sys-  
ivity, or to evaluate commercially  
uirements that have been identified.  
antages gained by transition to elec-  
Systems are generally implemented

either because the cost benefit analysis projects significant operational savings  
or because there are other business advantages to implementing electronic docu-  
ment management systems. In the best of cases, implementation of a computer  
based system will both reduce costs and provide additional advantages.

During the work flow analysis and the cost benefit analysis, it is a good idea  
to determine the overall objectives associated with implementing the system  
under consideration. The overall objectives should serve as a guide during sys-  
tem design and cost tradeoffs that may occur during the design and implementa-  
tion process. Examples of primary objectives on which implementation of new  
systems is based include the following:

**Productivity Improvement.** Electronic document management systems may  
be developed for use where the primary concern is improvement in productivity.  
In document processing applications, throughput is generally measured in the  
number of documents processed in a fixed amount of time. Use of electronic  
document management technology may yield improved productivity for specific  
steps within a given application.

**Reduction in Labor Cost.** Depending on the degree of automation imple-  
mented within a given system, it may be possible to reduce the number of indivi-  
duals involved in processing documents for a specific application. A cost benefit  
analysis is often used to justify the implementation of a new system based on  
labor savings achieved over a projected period of time. In applications where the  
document volume to be processed is always increasing, it may be cost effective  
to implement electronic document management technology instead of hiring  
additional staff.

**Reduction in Floor Space Cost.** Some applications involve periodic query  
of a large generally static paper data base. In these cases a cost benefit analysis  
that evaluates the costs for storage of the material versus the cost savings  
achieved by implementation of an electronic document management system can  
be performed to aid in the decision to change to computer based technology.  
Cost benefit analysis may also include consideration of the time required for  
manual retrieval of information from large paper data-bases versus retrieval  
from an electronic document management system.

**Improve Customer Response Time.** Time response is critical in many  
applications, including insurance claims processing, financial remittance form  
processing, loan application processing, and others. It is possible that  
implementation of an electronic management system will reduce the time  
required to support critical functions within a system and result in significant  
improvement in performance within specific applications. Electronic document  
retrieval can be faster than manual retrieval of files from paper storage, and  
documents can be routed to various locations electronically instead of manually.



**Data Security and Integrity.** Computer based systems can utilize available technology to restrict access to digitized document material. Various classes of users can be defined, each with different levels of access to document material which is categorized by the sensitivity of information contained within the document. Physical security of the facility containing the central digital database may be easier to achieve and less costly to implement when the database is stored on a single compute server than when it is stored in physical files of paper or microfilm. Misfiling of documents returned to the central file is also eliminated with the use of electronic document processing technology.

This list of overall system objectives is intended to be representative, and no attempt has been made to insure that the list is all-inclusive. It is important to note that each of these objectives drives requirements on portions of the electronic document management system. A system intended to reduce floor space may feature the highest digital density storage devices available (and possibly the most expensive), and the cost benefit analysis might indicate that the increased equipment cost required to obtain the latest technology with the smallest footprint might be worth it in the long run if the reduction of floor storage space is a significant system objective. A system that is being installed to reduce customer response time might feature the commercial networking equipment with the highest available bandwidth. Here again, the performance improvement may justify the additional investment in system capability.

## INITIAL SIZING OF DOCUMENT PROCESSING SYSTEMS

The first step in implementing or procuring an electronic document management system is an initial sizing and scoping of the system. This is often done during the cost benefit analysis, once the system objectives and functions have been identified, in order to obtain a rough cost of the system required to meet stated requirements and objectives. This section describes the sizing of a few systems, in some detail, to provide insight into the first stages of detailed system design.

There are several key considerations that dictate the type of architecture and the hardware content of electronic document management systems. These include the following:

**Duration of Document Storage On-line.** The time period for which document material must be retained on-line for query, retrieval, and processing (coupled with the volume of document material processed in the application) are the main drivers on the amount of digital on-line storage that must be provided.

**Daily Processing Volume.** The number of documents per day that must be processed is a major driver on system design. It drives requirements on scanning

based systems can utilize available document material. Various classes of levels of access to document material information contained within the containing the central digital database to implement when the database is it is stored in physical files of paper turned to the central file is also nt processing technology. intended to be representative, and no it is all-inclusive. It is important to quirements on portions of the elec- tem intended to reduce floor space age devices available (and possibly t analysis might indicate that the he latest technology with the small- un if the reduction of floor storage tem that is being installed to reduce commercial networking equipment gain, the performance improvement m capability.

## NG OF SING SYSTEMS

an electronic document management e system. This is often done during objectives and functions have been f the system required to meet stated scribes the sizing of a few systems, it stages of detailed system design.

dictate the type of architecture and nent management systems. These

*ine.* The time period for which for query, retrieval, and processing rial processed in the application) are -line storage that must be provided.

of documents per day that must be 1. It drives requirements on scanning

capacity, on-line storage, and communications speed between multiple points within the system (e.g., between optical disk storage and workstations).

**Workflow Considerations.** The number of different operations to be performed on documents, and the flow of documents between processing stages within the system drives the overall topology of the hardware and associated networks. It also drives the complexity of the software required to support the application.

At this point, the objectives for three different systems are described, and the impact of individual requirements on the sizing of the final system are illustrated. These cases are intended to be representative, and are not inclusive of all situations that will be encountered. They are included here to provide guidance in overall system sizing since similar calculations can be performed for a variety of applications in order to size the system to the point where a rough system cost can be estimated.

### Transaction Processing Systems—Example 1

Two different transaction processing applications will be reviewed here. The first system must be designed to operate five days/week, eight hours/day. It must be designed to accommodate a daily input rate of 50,000 8.5 x 11 inch document pages per day, single-sided. In this system, workflow analysis indicates that the acceptable time objective for completing several different processing steps on the input material is four days, and that each page of material is accessed on the average of four times during this time period. An additional requirement is for on-line access to all documents scanned for a period of one year after initial scanning. It is estimated that each page will be accessed once every 30 days during the one year retention period.

**Scanning Considerations.** Assuming a relatively high speed scanner can produce 3,000 pages per hour (including time for quality control and rescans), a total of 16.7 hours of scanning would be required. The system operations are constrained to eight hours/day. This means that the system designer could select two high speed scanners with some risk that the daily throughput requirement could be met. On the other hand, selection of three scanners would provide a comfortable margin of scanning performance while introducing additional costs for equipment and possibly another scanner operator.

**On-Line Storage Requirements.** There are two quite different on-line storage requirements for this system. The storage capacity must be designed to handle both the immediate material being processed (four days worth of data accessed on the average of four times during this period) and the archived material from prior processing (one year's worth of data accessed on the average of once every 30 days). In this scenario, it is likely that the system designer



would elect to maintain the immediate material being processed (the four days worth of data) on magnetic storage to provide the fastest possible access during the time period when it is being accessed most frequently. This choice would also provide the fastest possible transfer of scanned data into the system, since optical disk access and read/write times are slower than magnetic disk. The document material would be transferred to optical disk jukeboxes following completion of the immediate processing steps during the first four days that the document material is in the system.

With this design approach, the amount of on-line magnetic storage required for the application can be estimated. Assuming each scanned page would require 50 KBytes when compressed with an average compression factor of 10:1, the on-line magnetic storage required to store four days worth of data would be:

$$4 \text{ days} \times 50,000 \text{ pages/day} \times 50,000 \text{ bytes/page} = 10 \text{ GBytes}$$

The system designer will increase this amount of storage to accommodate uncertainty regarding the degree of compression achievable on the actual document material for the application, to allow some margin for error and processing periods that extend over the four day requirement, and to allow for on-line storage of other data required by the system (including system software, index data, associated catalog data, etc.).

The amount of on-line optical disk storage can be estimated in a similar manner. Here the requirement is to store one year's worth of data. In this example, there are 5 days/week  $\times$  52 weeks/year or 260 days of operations. Note that holidays are ignored in this example for conservatism in estimating storage requirements. The total amount of on-line storage required to accommodate one year's worth of data would be:

$$260 \text{ days} \times 50,000 \text{ pages/day} \times 50,000 \text{ bytes/page} = 650 \text{ GBytes}$$

If 12 inch WORM media is utilized, holding 4 GBytes/platter, 163 disks will be required to accommodate this data volume. The system designer will have to determine whether to hold all disks in a single jukebox with multiple drives or to distribute the data across two or more jukeboxes, based on the workflow analysis within the total system. Here again, the system designer will increase this amount to provide for significant design margin, as described for the magnetic storage case.

**Data Transfer Rates.** Four different data transfer rates must be considered in designing this system: (i) the transfer rate required to transfer compressed images into the system during the scanning process, (ii) the transfer rate required to support the 4 accesses per page during the first four days of processing, (iii) the transfer rate required to transfer the compressed images from magnetic to optical disk storage following the 4-day processing period, (iv) the transfer rate

l being processed (the four days the fastest possible access during st frequently. This choice would nned data into the system, since slower than magnetic disk. The ptical disk jukeboxes following uring the first four days that the

n-line magnetic storage required each scanned page would require compression factor of 10:1, the days worth of data would be:

bytes/page = 10 GBytes

ount of storage to accommodate n achievable on the actual docu-e margin for error and processing ent, and to allow for on-line stor-ding system software, index data,

an be estimated in a similar man-'s worth of data. In this example, days of operations. Note that holi-ism in estimating storage require-quired to accommodate one year's

bytes/page = 650 GBytes

g 4 GBytes/platter, 163 disks will The system designer will have to jukebox with multiple drives or to es, based on the workflow analy-system designer will increase this gin, as described for the magnetic

transfer rates must be considered ; required to transfer compressed ccess, (ii) the transfer rate required first four days of processing, (iii) pressed images from magnetic to ssing period, (iv) the transfer rate

required to support the retrieval of each page from long term storage on the average of every 30 days during the one year archival storage period.

It is assumed that image compression is performed either by the scanner or by the scanner server, so that the system input rate corresponds to the rate required to transfer compressed images into the system. The data rate at which image data will be entered into the system during scanning can be computed as follows:

$$(50,000 \text{ pages/day} \times 50,000 \text{ Bytes/page}) / (8 \text{ hr/day} \times 3,600 \text{ sec/hr}) \\ = 87 \text{ KBytes/second}$$

This is the same rate at which images must be transferred from magnetic to optical disk storage each day.

During the 4-day active processing period, every page will be accessed four times, or an average of once per day. If the system is processing 50,000 pages every day, there will be 200,000 pages on-line in magnetic storage at any given time. If each of these pages is accessed once per day, the data transfer rate will be:

$$200,000 \text{ pages} \times 50,000 \text{ bytes/page} / (8 \text{ hr/day} \times 3600 \text{ sec/hr}) \\ = 347 \text{ KBytes/sec}$$

Finally, the data transfer rate associated with accessing every page from optical disk storage an average of once every 30 days can be estimated. One year's worth of data will correspond to 260 days/year  $\times$  50,000 pages/day or 13 million pages. One-thirtieth of this volume is 433,000 pages. Thus, the volume of data to be transferred from optical disk every day is a total of

$$(433,000 \text{ pages} \times 50,000 \text{ Bytes/page}) / (8 \text{ hr/day} \times 3600 \text{ sec/hr}) \\ = 752 \text{ KBytes/sec}$$

For this example, the total data transfer rate requirement for the system can now be computed as the sum of the four individual data rates:

Scan transfer rate: 87 KBytes/sec

Document access during first four days: 347 KBytes/sec

Transfer from mag to optical disk: 87 KBytes/sec

Document access during one year retention: 752 KBytes/sec

Total Data Transfer Rate: 1.273 MBytes/sec

Converting from bytes to bits yields an average data transfer requirement of 10.2 Mbits/sec during an eight hour day.

The system designer must now select a network technology and design the overall topology of the network to support this volume of data transfer within the system. But there is still more work to do before completing the system design. Computer networks are generally rated by bandwidth, the total capacity



for data transfer measured in bits per second. However, there is a great deal of inefficiency encountered when transferring data across computer networks. Image data will be segmented into blocks or packets of bits. Each packet will also include extra bits required by network protocols. The number of bits per packet is generally based on system designs for transfer of alphanumeric data, and is not optimal for transfer of image data where large numbers of bits are involved. The system designer must account for these inefficiencies and increase the numbers computed above.

In addition to packet size inefficiency, most networks cannot operate close to their full bandwidth due to contention for network resources that occurs when multiple processors on the same network demand communications services at the same time. System designers must also take network contention into account when designing document management systems.

Finally, the computations presented above are *average* and not *peak* values. The designer must take into account the fact that the workload will not be spread evenly through an eight hour day, or even within the same few minutes in a distributed system. Systems must be designed to accommodate anticipated peak rates, not average rates, so that contention between tasks does not degrade system performance. The designer will incorporate another factor to translate average values into peak values before completing system design. Sophisticated queuing models may be used to model various alternative designs to determine the optimal network topology that will achieve minimum response time for all tasks performed using the system.

The example given above indicates that an average of at least 10.2 Mbits/sec of bandwidth would be required to support the application, based on the stated requirements. A network technology that supports data rates in excess of 10.2 Mbits/sec will be required to support this application, based on the considerations described above, the desire to retain a conservative margin in the design, and the desire to provide expansion capacity for the installed system. Current Ethernet networks based on twisted pair cabling provide bandwidths of 10 Mbits/sec, but the bandwidth is significantly degraded due to contention and packet protocol requirements. The system designer in this example would consider higher bandwidth network alternatives to support this application, or could consider a topology incorporating multiple Ethernet-based sub-networks. Detailed consideration of the workflow requirements, the most frequently utilized data transfer paths, and the overall topology of the installed system will dictate the final choice of network technology. This example illustrates the major impact of image data volume on communications links within electronic document management systems.

### Transaction Processing Systems—Example 2

This example is the same example as the first system, except that the data retention period is three years instead of one year and the access requirement on the

However, there is a great deal of data across computer networks, in packets of bits. Each packet will have its own header and trailer, and protocols. The number of bits per packet for transfer of alphanumeric data, where large numbers of bits are transferred, is small. This leads to these inefficiencies and increases the time to transfer data.

st networks cannot operate close to network resources that occurs when demand communications services take network contention into account ms.

are *average* and not *peak* values. That the workload will not be spread within the same few minutes in a distributed system to accommodate anticipated peak loads between tasks does not degrade system performance. Another factor to translate average workload into system design. Sophisticated alternative designs to determine the minimum response time for all

an average of at least 10.2 MBits/sec for the application, based on the stated supports data rates in excess of 10.2 application, based on the conservative margin in the design, for the installed system. Current cabling provide bandwidths of 10 Gbps, degraded due to contention and designer in this example would continue to support this application, or could use Ethernet-based sub-networks. Requirements, the most frequently utilized of the installed system will be 10 Gbps. This example illustrates the communications links within electronic

### —Example 2

st system, except that the data retention and the access requirement on the

long term storage database is once every 60 days instead of once every 30 days. All other requirements are the same as the first example (the system must be designed to operate five days/week, eight hours/day, must accommodate a daily input rate of 50,000 8.5 x 11 inch document pages per day, single-sided, the acceptable time objective for completing several different processing steps on the input material is four days, and each page of material is accessed on the average of four times during this time period).

**Scanning Considerations.** The scanning requirements are the same for this example as for the first example.

**On-Line Storage Requirements.** There are two quite different on-line storage requirements for this system, as in the first example. The storage capacity must be designed to handle both the immediate material being processed (four days worth of data accessed on the average of four times during this period) and the archived material from prior processing (three years worth of data accessed on the average of once every 60 days). The same division between magnetic storage and optical storage is assumed in this example. The volume of magnetic storage required for this example is the same as the first case (10 GBytes). Again in this case, the system designer would increase this amount of storage to accommodate uncertainty regarding the degree of compression achievable on the actual document material for the application, to allow some margin for error and processing periods that extend over the 4 day requirement, and to allow for on-line storage of other data required by the system (including system software, index data, associated catalog data, etc.).

The amount of on-line optical disk storage can be estimated as

$$3 \times 260 \text{ days} \times 50,000 \text{ pages/day} \times 50,000 \text{ bytes/page} = 1,950 \text{ GBytes}$$

If 12 inch WORM media is utilized, holding 4 GBytes/platter, 489 disks will be required to accommodate this data volume. This is clearly a case where a single jukebox will not suffice if the media capacity is 4 GBytes/disk. The designer will provide a sufficient number of jukeboxes to accommodate this volume, adding a comfortable margin. Operationally, documents would be distributed uniformly across multiple jukeboxes as soon as the long term storage material is transferred to optical disk at the start of system operations. This would provide for minimum contention in retrieving data from the optical disk jukeboxes, since multiple jukeboxes and drives could service data retrieval requests prior to the end of the first three years of operations. Note the increase in cost required to accommodate a three year, instead of one year, retention period.

**Data Transfer Rates.** The same four different data transfer rates must be considered in designing this system: (i) the transfer rate required to transfer compressed images into the system during the scanning process, (ii) the transfer rate required to support the four accesses per page during the first four days of processing, (iii) the transfer rate required to transfer the compressed images



from magnetic to optical disk storage following the four day processing period, (iv) the transfer rate required to support the retrieval of each page from long term storage on the average of every 30 days during the one year archival storage period. The first three data rates are the same as the first example. The data rate required to support retrieval of each page on the average of every 60 days from the three year archival storage database must be estimated.

One year's worth of data will correspond to 260 days/year x 50,000 pages/day or 13 million pages. Three year's worth of data will comprise 39 million pages. One-sixtieth of this volume is 650,000 pages. This is the volume of data to be transferred from optical disk every day, which is a total of

$$\begin{aligned} &650,000 \text{ pages} \times 50,000 \text{ Bytes/page} / (8 \text{ hr/day} \times 3600 \text{ sec/hr}) \\ &= 1.128 \text{ MBytes/sec} \end{aligned}$$

For this example, the total data transfer rate requirement for the system can now be computed as the sum of the four individual data rates:

Scan transfer rate: 87 KBytes/sec  
 Document access during first 4 days: 347 KBytes/sec  
 Transfer from mag to optical disk: 87 KBytes/sec  
 Document access during 1 year retention: 1128 KBytes/sec  
 Total Data Transfer Rate: 1.649 MBytes/sec

Converting from bytes to bits yields an average data transfer requirement of 13.2 Mbits/sec during an eight hour day for this example.

Several differences are apparent between the two examples presented here. In both cases, the system is intended primarily to support transaction processing, with the objective of completing processing of each input document within four working days. The requirement to retain information beyond the four day period turns out to be a major driver on system design. In both examples, the data transfer volume associated with infrequent access to the long term database dominates the total data transfer requirement. The design of transaction processing systems becomes significantly more complicated if the document material must be retained for access for any significant length of time beyond the standard processing time for the system.

The increase in data transfer volume between the two applications is approximately 30%. In both examples, the data transfer rates would stress all network technology except for the newer (higher cost) fiber optics based higher speed Local Area Networks (LAN's). In the first system, use of several lower speed LAN's might still meet the requirements. In the second example, the odds are that higher cost, higher bandwidth equipment will be required.

## Long Term Storage and Retrieval Example

This example will be typical of an application where digital image technology is being used to query and retrieve document material from a very large and relatively static database. The document database in this example is assumed to consist of 20 million individual pages. It is assumed that there is very minimal updating of this database on a daily or annual basis—only a few thousand additional pages are added to the database every year. In this example, the daily update to the database is not a major factor in sizing the system. Eight hour/day five day/week operation is assumed in this example, and it is assumed that each of the pages in the database is accessed on the average of once per year. In actuality, the vast majority of documents in this type of application is almost never accessed, and a small percentage of the documents is accessed more frequently, but the average is still computed over the whole database.

**Scanning Considerations.** In this application, the major scanning issue involves the initial conversion of the 20 million page database that must be performed before the system can come on-line for operational support. Again in this example, use of a relatively high speed scanner that can produce 3,000 pages per hour (including time for document preparation, quality control and rescans) is assumed. At that rate, it would take almost 7,000 hours to scan the full database, or 833 days assuming one shift operations. If four scanners are utilized for a full year, the entire database would be converted to digital image format at that rate. The cost of the database conversion turns out to be a major driving factor in the cost-benefit analysis for this system. In the earlier examples, two scanners were almost adequate to accommodate the daily transaction processing scanning requirements, but the system could come on-line the first day of operations. In this example, a year of operations involving expensive equipment that will be utilized only a short period of time and labor intensive document scanning is required before the full database is on-line and available.

**On-Line Storage Requirements.** In this example, the on-line storage requirement computation is straightforward. Again assuming 50 KBytes per compressed page, the total on-line storage requirement is 1,000 GBytes. Assuming 4 GBytes per 12 inch WORM platter, 250 disks are required to accommodate this data volume. Here again, the system designer will have to determine the data distribution across multiple jukeboxes (based on a detailed system design which is based on the workflow analysis within the total system) and will add margin to this computation to compensate for various factors in the estimation of total storage requirement and to allow for some future growth in the database.

owing the four day processing period, the retrieval of each page from long days during the one year archival re the same as the first example. The each page on the average of every 60 atabase must be estimated.

d to 260 days/year x 50,000 pages/day f data will comprise 39 million pages. ges. This is the volume of data to be ich is a total of

age / (8 hr/day x 3600sec/hr)  
ytes/sec

r rate requirement for the system can  
dividual data rates:

ys: 347 KBytes/sec  
: 87 KBytes/sec  
ention: 1128 KBytes/sec

MBytes/sec

n average data transfer requirement of  
r this example.

n the two examples presented here. In  
rily to support transaction processing,  
g of each input document within four  
nformation beyond the four day period  
esign. In both examples, the data trans-  
cess to the long term database domi-

The design of transaction processing  
licated if the document material must  
ngth of time beyond the standard pro-

etween the two applications is approxi-  
transfer rates would stress all network  
(cost) fiber optics based higher speed  
st system, use of several lower speed  
. In the second example, the odds are  
ent will be required.



**Data Transfer Rates.** In this example, the dominant factor in computing the data transfer rate is the retrieval of document pages from the database. If it is assumed that each page is retrieved on the average of once per year, the number of pages retrieved per hour can be computed as:

$$\begin{aligned} &20,000,000 \text{ pages/yr} / (8 \text{ hr/day} \times 5 \text{ days/week} \times 52 \text{ weeks/yr}) \\ &= 9,615 \text{ pages/hr} \end{aligned}$$

The data transfer rate, assuming 50 KBytes/page turns out to be 133 KBits/second for an eight hour shift. This transfer rate is well below those computed for the transaction processing system described in the earlier examples. Even though this is a major system that would obviously involve several operators working simultaneously on the database to achieve the 9,615 pages/day access rate, the data transfer requirements in this system can probably be achieved using less expensive, lower bandwidth technology than appeared to be required by the transaction processing systems described above.

Here again, detailed system design would consider average versus peak communications rates and incorporate additional design margin and conservatism to insure lack of contention between operators and acceptable system response time for the operators using the system.

#### **An Alternate Approach to Backfile Capture**

The last example illustrates one case where capture of the existing document database from either paper or film files (microfilm aperture cards, microfiche) would be a major cost driver in conversion of an operation to an electronic document management system. The example assumed that the entire existing document file had to be converted into electronic images before the system was declared operational. The set of existing document records that are either on paper, film, or both, that exists in a particular application at the time that the operation converts to digital processing is often referred to as the *backfile*.

There is an operational approach that can be used to minimize the costs of backfile capture. The approach involves a phased transition of the document file to digital format. When the new electronic document management system comes on-line, it is designed to convert all new incoming material to image format, so that all "day forward" new material entered into the system will always be digitized. Whenever a user calls for a document that is located in the backfile, the document is retrieved from paper or film records and converted to digital format and properly indexed at that point. This process continues during operation of the new system, and over a period of time the backfile is slowly converted to digital format. In many applications, some backfile documents are never requested again, while some documents in the backfile may be retrieved several times during normal operations. In many cases, the backfile documents may never be converted to electronic format.

e, the dominant factor in computing the  
ment pages from the database. If it is  
e average of once per year, the number  
ted as:

$\times 5 \text{ days/week} \times 52 \text{ weeks/yr}$   
ages/hr

ytes/page turns out to be 133 KBits/sec-  
rate is well below those computed for  
ed in the earlier examples. Even though  
sly involve several operators working  
e the 9,615 pages/day access rate, the  
can probably be achieved using less  
than appeared to be required by the  
above.

ould consider average versus peak com-  
nal design margin and conservatism to  
ators and acceptable system response

### Capture

ere capture of the existing document  
(microfilm aperture cards, microfiche)  
n of an operation to an electronic docu-  
assumed that the entire existing docu-  
tronic images before the system was  
; document records that are either on  
icular application at the time that the  
often referred to as the *backfile*.

can be used to minimize the costs of  
phased transition of the document file  
: document management system comes  
incoming material to image format, so  
ed into the system will always be digi-  
ent that is located in the backfile, the  
records and converted to digital format  
process continues during operation of  
ne the backfile is slowly converted to  
ome backfile documents are never  
the backfile may be retrieved several  
y cases, the backfile documents may

This approach minimizes the effort required for backfile conversion. The scanning, indexing, and quality control operations are performed only on backfile documents that are still in active use, and the total amount of labor and equipment required for backfile capture may be significantly less than that required to convert the entire backfile.

The disadvantages of this approach may include the fact that the backfile material may be degrading with time, and "on demand" backfile conversion may eventually result in a significant loss of portions of the original material. The "on demand" approach to backfile conversion may also operate more slowly than a routine batch production scanning operation that is directed toward systematic capture of the entire backfile. In the "on demand" approach, individual documents must be located on paper or film, retrieved, and scanned individually. Evaluation of the "on demand" backfile capture approach for a specific application is also performed as part of the overall system design study.

## ADDITIONAL SYSTEM DESIGN ISSUES

This section contains a summary listing of issues that are generally addressed during the system design or system procurement process. This listing is not intended to be a complete checklist, but it does include some major considerations that are generally dealt with when an operational activity is phased over to an electronic document management system.

### Spatial Resolution

The subject of spatial resolution of scanned imagery has been dealt with at some length in Chapter 2, and the impact of scanning resolution on image quality is also discussed in the Advanced Topics section of the text. Choice of scanning resolution ultimately represents a tradeoff between communications time for documents to be transferred between points of the system, computation time required for compression/decompression, and the quality of the scanned imagery and legibility of the information contained within the documents.

Selection of the scanning resolution affects other system features as well. One example is the appearance of document material on display screens and the degree to which document information can share display space with other information (such as pull down menus and other user interface or control information). Figure 3.5 showed three different ways in which a 1,700 x 2,200 pixel high resolution binary display can be used for document and user interface information. If the document material has been scanned at 300 dpi, an image of an 8.5 x 11 inch page will be too large to be displayed on a 1,700 x 2,200 pixel screen. Figure 3.5 showed two different display modes for imagery scaled to 150 dpi using hardware or software. If the 8.5 x 11 inch document was scanned at 200 dpi, then the 1,700 x 2,200 pixel display was adequate to display the full



image in portrait mode, but there was no room left on the display for user interface information.

Once the scanning resolution has been selected, additional decisions must be made regarding the methods used to display scanned images on workstations, and methods used to produce hard copy from scanned imagery. In most systems, hardware or software scaling is provided at various points in the system to accommodate differences between scanning resolution and output resolution. Many systems have been designed to capture document material at the highest resolution deemed appropriate for the application, and then scale the information to lower resolution for specific operations at various points in the processing.

## Document Indexing

Effective document retrieval, and effective query of document databases, require that an appropriate scheme for indexing be defined for every electronic document management system. Most systems either provide indexing, query, and retrieval capability or utilize commercially available database management systems to support these functions.

The application dictates the type and degree of document indexing required. In some systems, a unique identifier assigned to every scanned page is adequate. Check processing systems and transaction processing systems often involve an index scheme utilizing only the discrete unique identifier of every image scanned. Most systems, however, require a more complex indexing scheme.

Indexing schemes often reflect the logical structure of the document database. Digitized images are organized into individual pages, documents that consist of one or more pages, files that may contain one or more documents, etc. Material can be retrieved based on these indices at any level (page, document, or file). Some indexing schemes also retain content information as part of the database index, so that searches can be initiated for all documents that contain specific information. Indexing schemes may also involve a time reference base, so that queries can be made for all documents that meet a specific search criteria within a specified time period.

Index information is most often associated with document material at the point of initial entry of information into the system—the scanning operation. During scanning, or in a separate step just after scanning, data can be entered by operators or automatically by the system that provides the index keys that will support the required types of database query and retrieval. Automated indexing, involving the use of bar code recognition or optical character recognition technology may be used, with or without manual quality control review of assigned index values. These technologies are described further in the Advanced Topics section of this text.

## Security and Data Access Issues

One advantage of the use of electronic document management systems versus paper or film files is the degree of data access control provided by computer technology. It is possible to define several types of users based on the degree of access provided to the document database. Some individuals may be restricted to query and retrieval of specific classes of documents, for example. It is possible to partition the digital document database into subsets or segments and to control access by individuals to each of the subsets or segments. It is also possible to establish an audit trail, maintaining a record of access to each document, including the individual, date and time of access, etc. It is also easy to prevent public access to portions or all of a document database if desired.

Data access is another issue that should be considered during the system design or procurement process. Early decisions are required regarding the access restrictions to be applied to specific classes of documents, the various types of operator access required to perform the functions within the application, and the levels of security control for the overall system.

## Centralized Versus Distributed Functions

One of the results of the overall workflow analysis that occurs prior to system selection or design is a series of decisions regarding the approach to allocation of several system functions. General decisions are made regarding centralization or distribution of specific operations within the overall system. A *centralized* function is performed at one location within the overall system, and a *distributed* function can be performed at more than one location.

One example of a function for which this decision is required is document scanning. In one application, document entry into the system may be a well-defined production procedure. In an insurance application, for example, all claims may arrive at a central mailroom, and the scanning, document indexing, and quality control functions can best be performed in that location using dedicated scanning equipment, associated workstations, and support software. In a different application, it may be that document material arrives at various points in the system (different departments or branch office operations of one organization), and it may be more efficient to input document material into the system at more than one physical location. If that is the case, all scanning locations will require the same capability in terms of scanning equipment, workstations, and support software.

The location of the document database represents another design decision. For some applications, a centralized database serving a distributed operation may be desirable. For other applications, it may be desirable to divide the data-



base between several different locations, each with a specialized set of tasks to perform. The centralized database concept provides a single point of access and distribution of all document data within the system. The distributed database concept may or may not require that all locations within a given system have the capability of accessing any document within the system, wherever it is stored. Here again, workflow analysis, tradeoffs between communications bandwidth and equipment expense, and cost tradeoffs generally indicate the appropriate solution for a given application.

Generation of hard copy output is another function that can be centralized or distributed. A centralized hard copy generation capability provides better control of information distribution via hard copy, and that degree of control may be desirable in a specific application. Distributed hard copy generation provides efficient response to output requirements at individual locations within a system, but it may result in a lack of control of information dissemination. Analysis of workflow, hard copy volume requirements, cost, and communications link tradeoffs is required to determine the optimal manner in which hard copy output will be provided within a given system.

### User Interface Considerations

Workflow analysis will generally determine the allocation of individual functions to multiple operator positions within a given system. Some systems may involve use of a centralized database with multiple operators all performing the same functions. More complex systems may involve several different classes of operator stations, with document material flowing from the entry point in the system through a series of processing steps that are performed in different parts of the system, typically on different workstations. In the latter system, different software is required to support each function assigned to a different operator position, while for the first example the same software is present in all locations.

Software generally supports the operator interaction with the system. Individual tasks or operations performed within a given system on document material will require different types of access to document information, and software must also support all operator actions that will be taken based on review of document content at each point in the system. In some simple systems, for example, the only requirement may be for operator query, retrieval, review, and on-demand printing of document information. More complex systems may involve several stages of processing. For example, in a claims processing operation, it may be necessary to support the following set of operator functions, each of which may be performed at physically different locations within the system:

- i) Document scanning, indexing, quality control and rescan
- ii) Initial review of claim for completeness and accuracy with generation of correspondence as required, routing of claim to appropriate processing point

th a specialized set of tasks to  
es a single point of access and  
tem. The distributed database  
within a given system have the  
system, wherever it is stored.  
n communications bandwidth  
erally indicate the appropriate

tion that can be centralized or  
pability provides better control  
hat degree of control may be  
ard copy generation provides  
dual locations within a system,  
ion dissemination. Analysis of  
nd communications link trade-  
in which hard copy output will

allocation of individual func-  
en system. Some systems may  
le operators all performing the  
olve several different classes of  
ag from the entry point in the  
re performed in different parts  
.. In the latter system, different  
signed to a different operator  
ware is present in all locations.  
interaction with the system.

a given system on document  
ocument information, and soft-  
ill be taken based on review of

In some simple systems, for  
or query, retrieval, review, and  
More complex systems may  
; in a claims processing opera-  
set of operator functions, each  
t locations within the system:

ontrol and rescan  
accuracy with generation of cor-  
appropriate processing point

- iii) Review of claim for validity, with generation of correspondence for claims not valid
- iv) Determination of compensation, including access and retrieval of standard data files in addition to the claim itself
- v) Supervisor review of actions taken, including initiation of correspondence, printing of check reimbursement, entry to/from accounting database
- vi) Overall system management, including workflow monitoring, reassignment of functions to operators, monitoring and assessment of operator performance, generation of management reports

Each of these individual functions will require support from a display and software that provides the operator with the capabilities to perform each of the functions listed above. Design of an interface supporting each of these functions involves many complex decisions, including a consideration of its ergonomic impact. These functions will generally be performed repetitively every day by the same operator, so issues of operator ease of use, operator fatigue, and use of interactive equipment to control specific functions (e.g., a mouse) are all involved.

### **Document Presentation to the Operator**

There are many ways of presenting document information to the operator using display equipment. If documents within an application consist of a single page only, the issues are simplified. In cases where documents may consist of more than one page, and the operator will require access to multiple pages within the document, various methods have been developed that make it possible for the operator to view the total document in order to select the page or pages of interest. Figure 6.1 shows two different methods.

The left side of Figure 6.1 shows a display containing an image of every page of a document presented to the operator in a single display at highly reduced resolution. The resolution must be adequate for the operator to determine the general content of each page. The operator then utilizes software (and possibly some interactive hardware) to select the pages required to complete the assigned task. This is a good way to display documents that may consist of a mixture of standardized forms and associated document material that is less structured (such as correspondence, receipts, and copies of supporting material).

A second method of displaying a document is shown on the right side of Figure 6.1. Here the individual pages are shown in a "stacked" format, with the edges of each successive page visible to the operator. If enough material from each page in the document is visible to the operator, it should be possible to select individual pages. It may also be possible to "browse" the document at full screen resolution (in a forward or reverse direction) using this type of interface.



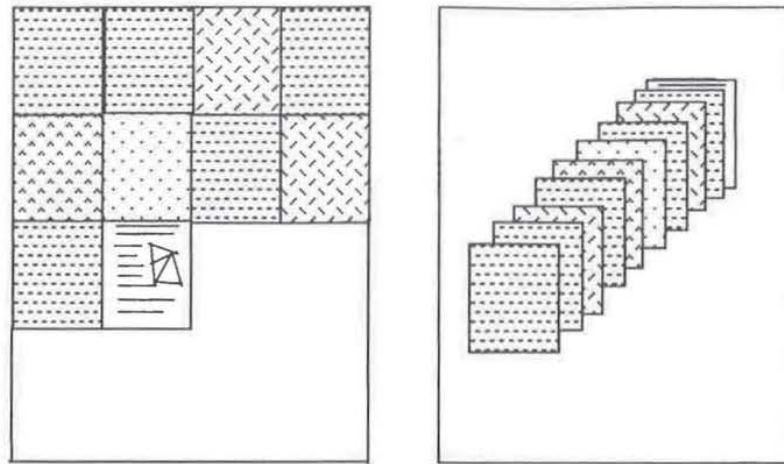


Figure 6.1 Two different examples of document display for operator review.

### The Impact of the Open Systems Philosophy

In the early days of the computer industry (the 1950s through the 1970s), most computer vendors sold proprietary hardware and software solutions to customer problems. Purchase of a particular vendor's systems involved an internal commitment to long term utilization of that vendor's products. Transition from one vendor to another required a significant investment in terms of training on new equipment, new operating systems, new communications protocols, and sometimes new programming languages. Peripheral hardware procured to operate on one vendor's equipment would often not operate as a peripheral on any other vendor's equipment because the interface and communications protocols were proprietary.

In the 1980s, customer pressure and the emergence of industry-wide standards caused a significant change in the business environment within the computer industry. Standards started governing the most widely utilized technology. Standardized communications protocols emerged, enabling data transfer between equipment manufactured by different vendors. Operating system standards were driven by the widespread acceptance of the Unix operating system. Standardized display protocols also emerged. By 1990, the market was being driven by standardization efforts.

The International Standards Organization established a set of standards governing communications between processors, called the Open Systems Interconnection (OSI) standard. This standard defined seven different layers of interaction that could occur between disparate processors. The lowest layer is the physical linkage. Higher layers contain components at increasing levels of

Syste

abstra  
ized c

Th  
stand  
benef  
and it  
softw  
addre  
impac

i)

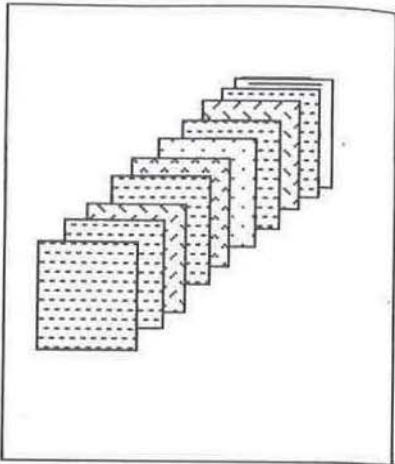
ii)

iii)

iv)

v)

Use  
standa  
port a



ent display for operator review.

## is Philosophy

(the 1950s through the 1970s), most e and software solutions to customer s systems involved an internal com- andor's products. Transition from one /estment in terms of training on new m- munications protocols, and some- eral hardware procured to operate on p- operate as a peripheral on any other and communications protocols were

e emergence of industry-wide stan- is- iness environment within the com- the most widely utilized technology. i emerged, enabling data transfer ent vendors. Operating system stan- tance of the Unix operating system. ed. By 1990, the market was being

1 established a set of standards gov- ssors, called the Open Systems urd defined seven different layers of rate processors. The lowest layer is components at increasing levels of

abstraction, including applications software. The emergence of these standard- ized definitions has led to development of so-called Open Systems architecture.

The result of the standardization embodied in the OSI structure and other standards that have emerged based on the same general philosophy has been beneficial for the customer making the investment in hardware and software, and it has increased the competitive pressure within the computer hardware and software vendor community. It has also broadened the market that can be addressed by individual components manufactured by the vendors. The main impact of standards can be summarized as follows:

- i) Peripheral devices now communicate with processors via standardized hardware and software interfaces. This means that a standards-compliant peripheral device (such as a scanner or a disk drive) can operate on processors manufactured by many vendors. Still more important, the peripheral device will probably also work on the next generation proces- sor as well, so that users can take advantage of improved processor per- formance without losing their investment in peripheral equipment.
- ii) Software language standards and operating system standards means that software that operates on one processor will also work on processors from other manufacturers. Here again, the user's investment in software development and training is protected, since the software and operating system can migrate to newer generations of hardware with minimal mod- ification effort.
- iii) Standardized communications protocols make it possible to connect sys- tems for several different vendors within the same networked distributed system. Systems that consist of networked configurations containing processors and equipment from multiple vendors are often called "het- erogeneous systems."
- iv) Standardized display protocols make it possible to display images and other information on equipment of varying resolution and capability man- ufactured by more than one vendor. The investment in software and user training is again protected, and systems can evolve into newer, more capa- ble equipment without major investments in new software development.
- v) In the document processing world, standards have evolved that make it possible to run the same applications on more than one vendor's hard- ware. One example is the use of the CCITT image compression standard for binary image compression. As seen earlier, almost all board level products used in electronic document management systems support CCITT compression and decompression functions.

Use of Open System methodology (in other words, compliance with industry standards) makes it possible to procure the latest most capable hardware to sup- port a specific task. Users are assured that they can make the most cost-effective



purchasing decisions at any time, because much of the development investment is protected by standards compliance.

The Open Systems approach is especially relevant to newly emerging technology, such as electronic document management. Most systems used to process documents are relatively new. As the technology matures, it can be anticipated that the variety and type of applications for this technology will expand. In addition, many installations will initially commit to conversion of only a portion of their operation to computer based technology. As more applications migrate to the system, it will be forced to expand or be replicated. Open Systems compliance will insure that systems can migrate to the latest available hardware and achieve the benefit of improving price-performance ratios that characterize the computing industry.

When evaluating a system, it is important to determine the degree of standards compliance of any system procured from any vendor to guarantee maximum flexibility in the future. Similarly, those designing and developing new systems can benefit from standards compliance since it will simplify the development effort and insure graceful expansion capability in the future.

## Exam

The number and  
agement system:  
benefits of elect  
tions. The major  
ment technolog  
material. More r  
applied to impr  
ented applicatic  
systems have en  
tions can be cla  
systems have b  
combine attribut

This chapter  
Several exampl  
examples of the  
lier in this text t

## STC

The first system  
were storage al

th of the development investment

relevant to newly emerging technology. Most systems used to process technology matures, it can be anticipated that technology will expand. In addition, conversion of only a portion of As more applications migrate to complicated, Open Systems comply with the latest available hardware and performance ratios that characterize the

to determine the degree of standard any vendor to guarantee maximum designing and developing new since it will simplify the development ability in the future.



## Examples of Electronic Document Management Systems

The number and variety of applications addressed by electronic document management systems continues to increase as the technology develops and the cost benefits of electronic imaging become increasingly clear for individual applications. The majority of the first systems utilizing electronic document management technology were used for archival storage and retrieval of document material. More recently, electronic document management techniques have been applied to improve productivity and reduce processing time in transaction-oriented applications. *Storage and retrieval systems* and *transaction processing systems* have emerged as the two broad categories within which most applications can be classified. Within these two classifications, many different types of systems have been developed to address specific functions, and some systems combine attributes of both categories.

This chapter describes the characteristics of the two categories of systems. Several examples of real installed systems are described, providing concrete examples of the application of the components and technologies described earlier in this text to real world situations and applications.

### STORAGE AND RETRIEVAL SYSTEMS

The first systems to incorporate electronic document management techniques were storage and retrieval systems. These systems developed during the late



1970s and early 1980s, in parallel with the introduction of the first commercially available optical disk storage devices and commercial scanners. Optical disk technology offered the promise of long term storage capability at a reasonable cost per bit, providing a competitive alternative to micrographics systems with all the advantages of digital technology discussed earlier in this text.

The main application for storage and retrieval systems involves query and retrieval of documents from a large central database by a distributed set of users. The central database contains a collection of documents that do not change over long periods of time, although periodic updating of the central database may occur. Examples of systems of this type include storage of large volumes of personnel records and storage of reference documents (e.g., patents) in Government applications. Documents are indexed by search indices relevant to the specific application, and these systems generally feature a limited number of user interfaces or may support a single user interface.

Implementation of a storage and retrieval system generally involves scanning of a large backfile of existing document material before the system first comes on line for general use. Scanning of hundreds of thousands of pages of material from paper or film sources is not uncommon when a new system is installed. Care must be taken to index the scanned backfile to enable efficient query and retrieval of the document material. Re-indexing a scanned backfile of documents is a very inefficient way to correct an oversight in the initial design of the search keys for a particular application.

User interfaces for these systems support query, retrieval, and display of document material, and these systems generally provide for generation of hard copy printout of documents that have been retrieved from the central database.

## TRANSACTION PROCESSING SYSTEMS

Transaction processing systems are generally designed to handle applications driven by a daily or periodic flow of document material. Most transaction processing systems are based on incoming forms or correspondence that must be processed in a timely manner. Examples of applications supported by electronic document transaction processing systems include insurance claims processing, processing of applications (driver's license, auto registration, employment, patent applications, etc.), and processing of financial documents (checks, remittance slips, stock certificates, etc.). The applications are generally characterized by a requirement to process information within a specific time period. They are often forms-driven, although many transaction processing systems also support processing of correspondence and material not provided on specific forms.

Tr  
main  
tains  
data  
off li  
syste  
onda  
The  
rent  
cessi  
infor  
displ

Tr  
proce  
cated  
the v  
passi  
ated  
differ  
ment

Th  
of sc  
ity, s  
tus c  
mane  
of in  
proce

This  
docu  
tems  
broa  
appli  
O:  
desig  
rate  
Com  
cessi

introduction of the first commercially commercial scanners. Optical disk storage capability at a reasonable price relative to micrographics systems with systems discussed earlier in this text.

Retrieval systems involves query and retrieval by a distributed set of users. Documents that do not change over time, such as the central database may include storage of large volumes of permanent records (e.g., patents) in Government records indices relevant to the specific user and a limited number of user inter-

A system generally involves scanning material before the system first comes into use of thousands of pages of material and when a new system is installed. It is a file to enable efficient query and retrieval of a scanned backfile of documents and it in the initial design of the search

query, retrieval, and display of documents to provide for generation of hard copy from the central database.

## SING SYSTEMS

designed to handle applications of material. Most transaction processing or correspondence that must be applications supported by electronic records include insurance claims processing, auto registration, employment, financial documents (checks, remittances) are generally characterized by a specific time period. They are processing systems also support provided on specific forms.

Transaction processing systems generally do not involve development and maintenance of a large static database. A database may be maintained that contains a record of all material processed within a specific time period, but the database is generally dynamic—older material is either discarded or moved to off line storage on a periodic basis. The main focus of transaction processing systems is expedited handling of the most recent incoming data, with a secondary focus on the ability to locate historical records of earlier transactions. The systems are generally designed to provide essentially instant access to current material and are optimized for rapid response times when retrieving or processing current information. They may also provide access to older historical information, and the response time is generally much longer for retrieval and display of historical information than is associated with retrieval of current data.

Transaction processing systems are often designed to support many different processing steps within a single system. Individual workstations may be dedicated to specific functions within the application. These systems may support the *workflow* concept; documents may flow sequentially through the system, passing automatically from workstation to workstation as the operations associated with the information processing are completed. The user interface may be different on different workstations within the same system, reflecting the assignment of individual functions to specific workstations.

The system may be designed to incorporate automatic forwarding and routing of scanned document material. It may include administrative reporting capability, so that management can monitor the flow of information and processing status on a periodic basis. Typical examples of statistics that may be used for management of the processing performed by these systems include the number of incoming forms per day, the number of pending activities at each stage of the process, and the number of completed functions for a given time period.

## OTHER SYSTEM CATEGORIES

This introductory section has described the two broadest categories of electronic document management systems. The distinction is not perfect, and many systems include elements of both types of systems as described above. Within these broad categories, specific types of systems have emerged that address individual applications areas or processing of specific types of information.

One example of a significant sub-category that has evolved are systems designed to process and maintain engineering drawings. These systems incorporate specialized equipment to accommodate large drawings, aperture cards, and Computer Aided Design (CAD) files and include can include transaction processing functions (incorporation of new drawings, automated routing of new



designs for approval, etc.) and archival storage and retrieval functions (serving as the repository for storage, retrieval and display of all design material for a particular activity).

Another example of a sub-category are systems designed for processing of financial documents, such as check processing systems. These systems are designed to scan incoming checks, may include character recognition or hand keying to extract amounts from checks, may support computation of trial balances, and may provide the capability to store check images on optical disk media for long term archival storage and retrieval purposes.

The remainder of this chapter contains summaries of several different types of electronic document management systems. Each of the systems described is a real system, in actual operation within the application that is described. For each system, a description of the application and the system architecture that was developed to support the application is described. This chapter is designed to show applications of the technology described in the earlier chapters to real world problems and applications. The systems described in this chapter are systems that involve processing of relatively large volumes of information, and they serve to illustrate the fact that the technology has evolved to the point where major applications problems can be addressed in a cost effective manner using electronic imaging systems.

### NATIONAL ARCHIVES AND RECORDS ADMINISTRATION OPTICAL DIGITAL IMAGE STORAGE SYSTEM (NARA ODISS)

As the central repository for permanently valuable records of the federal government, the United States National Archives and Records Administration has an ongoing responsibility to evaluate new technology for potential incorporation within its operations. In 1984, NARA's Archival Research and Evaluation Staff recommended that a pilot system be installed to evaluate the use of digital image technology and optical disk storage technology for long term storage and retrieval of document material. A specific subset of the NARA document files was recommended for use with the pilot system to provide a detailed evaluation of the operational factors associated with the use of digital image technology in a well-defined storage and retrieval application. A procurement was initiated in 1985, with contract award in 1986 and installation and operation and evaluation of results occurring in the 1988-1989 time period. This unique pilot program included evaluation of several different technology components that were commercially available in that timeframe, and quantitative performance results for document scanning, retrieval and other functions were defined and documented. A final report<sup>1</sup> that describes the project approach, the system and subsystems

Exam

that w  
of NA  
imple  
produ  
tion, t  
the fir  
The

i)

ii)

iii)

iv)

v)

vi)

Th  
tectur  
devel

The

The (c  
the c  
Pensi  
400 r  
A sul  
Tenn  
War  
ment  
other  
requi  
each  
slips.  
spon

ge and retrieval functions (serving display of all design material for a

systems designed for processing of sing systems. These systems are ude character recognition or hand support computation of trial bal- ore check images on optical disk val purposes.

maries of several different types of ach of the systems described is a lication that is described. For each the system architecture that was ribed. This chapter is designed to ed in the earlier chapters to real s described in this chapter are sys- e volumes of information, and they y has evolved to the point where d in a cost effective manner using

## AND RECORDS AL DIGITAL IMAGE NARA ODISS)

able records of the federal govern- and Records Administration has an nology for potential incorporation iival Research and Evaluation Staff to evaluate the use of digital image ology for long term storage and abset of the NARA document files em to provide a detailed evaluation use of digital image technology in on. A procurement was initiated in lation and operation and evaluation period. This unique pilot program nology components that were com- quantitative performance results for ions were defined and documented. proach, the system and subsystems

that were implemented, and quantitative results of operational testing on a subset of NARA's records is available. It is one of the most complete summaries of implementation and operation of a digital image storage and retrieval system yet produced. This section describes the application selected by NARA for evaluation, the ODISS system and its components, and some of the results contained in the final report.

The goals for the ODISS system are summarized in Reference 1 as follows:

- i) To establish the feasibility, costs, and benefits of converting paper and microform documents to optical digital media and to assess document input speeds required to accomplish conversion in an operational environment
- ii) To determine the optimal scanning density for documents consistent with producing legible images while minimizing storage requirements
- iii) To assess the storage capacity of the system and media in terms of storage cost and efficiency
- iv) To evaluate system capability to automatically retrieve stored optical images using electro-mechanical devices
- v) To determine the suitability of creating printed document images from digital data
- vi) To determine staff and public reaction to and acceptance of an image retrieval system as opposed to the paper and microfilm currently used for reference

These objectives drove the selection of system components, the overall architectural design of the ODISS system, and the design of the custom software developed to support the application.

### The Application

The General Reference Service Branch of NARA is responsible for maintaining the compiled military service records (CMSR's) of the United States, and the Pension and Military Service Records Section supports approximately 300 to 400 requests from the general public per day relating to military service records. A subset of the CMSR was selected for use as a pilot application for ODISS, the Tennessee Confederate CMSR records containing documents from the Civil War era. These records consist of a jacket for each individual and a set of documents enclosed in the jacket. The jacket contains the name of the individual and other relevant information, and scanning of the jacket for each individual was required. The scanned jacket image became the first image of the file created for each individual. The jacket contents could include reference cards, reference slips, envelopes, or folders containing additional documents, and general correspondence. The age of the material, and variation in paper size and weight, the



different writing instruments used to record information, and the general condition of the document material (much of it stained, discolored, folded, or damaged) provided an excellent sample set of material with which to evaluate modern electronic scanning and image processing technology.

The Tennessee CMSR records are filed in boxes, with each box containing approximately 150 to 180 soldiers' records. The paper files are maintained by regiment, with soldiers' files stored alphabetically by name. Physical search and access is provided by the NARA staff. A microfilm record of this material also exists, which provided a good comparison between the document preparation required for film versus digital scanning and other cost and efficiency comparisons. The results from the digital scanning process indicated that 53,783 individual files were contained in the Tennessee CMSR material, and 220,713 images were produced from this material. Five 12 inch optical disks were required to store the material; the paper document material occupies 80 cubic feet in the NARA archives.

Figure 7.1 shows one example of a scanned jacket from the Tennessee CMSR shown on a high resolution display that also displays user interface menus and

The figure displays two side-by-side images. On the left is a scanned paper record card for James Bond, Co. D, 7th (Duckworth's) Tennessee Cavalry. The card is handwritten and includes a list of numbers (48301417, 1670, 30007716) and a section for 'Number of medical cards' and 'Number of personal papers'. On the right is a digital user interface menu for the National Archives Optical Digital Image Storage System. The menu lists functions F1 through F10, including HELP, Display Page No., Zoom Page 2x, Page Rotate, Reverse video, Window Functions, Display at Original Resolution, Print, Print Screen, and CANCEL/SEARCH. It also includes navigation keys PgUp and PgDn, a note about Shift F1, and a page indicator FCM 00010790 Page 0001 of 0005. Below the menu is a data entry form for James Bond, showing his rank as Lieutenant, company as D, and regiment as 7th (Duckworth's) Cavalry.

**National Archives  
Optical Digital Image Storage System  
Staff Workstation**

F1 - HELP  
F2 - Display Page No. 0001  
F3 - Zoom Page 2x  
F4 - Page Rotate, Reverse video  
F5 - Window Functions  
F7 - Display at Original Resolution  
F8 - Print  
F9 - Print Screen  
F10 - CANCEL/SEARCH

PgUp - Display Next Page  
PgDn - Display Previous Page

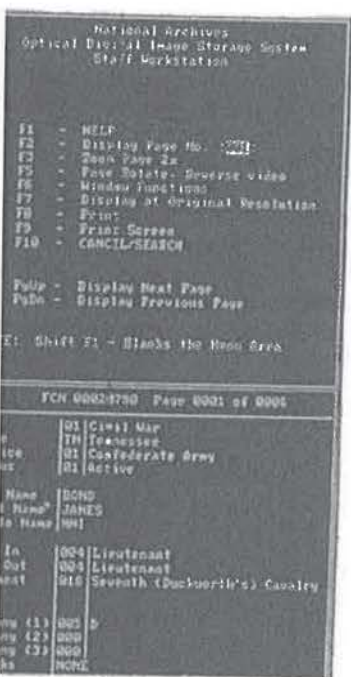
NOTE: Shift F1 - Blanks the Menu Area

FCM 00010790 Page 0001 of 0005

War	01 Civil War
State	10 Tennessee
Service	01 Confederate Army
Status	01 Active
Last Name	BOND
First Name	JAMES
Middle Name	WILLIAM
Rank In	004 Lieutenant
Rank Out	004 Lieutenant
Regiment	010 Seventh (Duckworth's) Cavalry
Company (1)	005 D
Company (2)	000
Company (3)	000
Remarks	NONE

Figure 7.1 Scanned jacket image from the Tennessee CMSR records. From Reference 1.

acket from the Tennessee CMSR  
displays user interface menus and



The system was designed to support several main functions. The first function was scanning, indexing, and quality control of the documents selected for this pilot study. A record was kept of the level of effort required for document preparation and scanning, and statistics were maintained relating to the percentage of scanned images that required rescanning for a variety of reasons. The second function was support of public requests for data by NARA staff members. Workstations with specially designed software supported query of the document database, retrieval of scanned images, and generation of hard copy output of specific documents. A third function involved placement of a search terminal with digital image capability in the NARA public search room and support of a remote terminal at the Tennessee Archives that enabled search of the indices of the documents but did not include image display capability. Studies were made of the operation of each of these functions and the advantages and disadvantages associated with use of digital imagery instead of paper documents or microfilm. These are described in Reference 1.

CMR records. From Reference 1.



## The ODISS System

Figure 7.3 shows the overall ODISS system as installed at NARA. Overall control of all operations, data flow, and routing within the system was performed by a centralized system manager, a microprocessor on which a commercially available database management system was resident. The bus architecture was based on Multibus, and direct RS-422 connections between the system manager and the workstation servers were utilized. This design was selected to minimize delays associated with transfer of data over slower commercially standardized networks that were not optimized for image data transfer.

A capture server microprocessor and custom software supported several different scanners. One of the major objectives of the ODISS pilot system was evaluation of the ability of commercially available scanning equipment to handle fragile documents, documents of different sizes and paper weight, and documents with varying image quality. Four separate scanners were installed within ODISS to support this objective. A high speed paper scanner was used for production or volume scanning of the majority of document material. Clear polyester sleeves were used to enclose the more fragile documents to prevent damage during high speed scanning operations, which worked satisfactorily with the high-speed scanner. The high-speed scanner operated at 200 dpi, scanning and compressing up to 26 pages per minute, double-sided. Custom software supported operator control of scanning operations and indexing of individual documents.

Two lower-speed paper scanners were also provided. A flat bed paper scanner was provided for scanning of document material that was too fragile to be scanned on the high-speed production unit. A third paper scanner, also a flat bed unit, produced gray scale imagery (8 bits per pixel) and was provided to support an image enhancement system incorporating various image processing and enhancement algorithms. This image processing subsystem was used to experiment with application of image enhancement techniques to poor quality documents. Different image processing algorithms were applied to scanned 8 bit/pixel imagery to improve the quality of scanned material whose original quality was poor.

A custom film scanner was included in the ODISS to evaluate the quality of document material scanned from a variety of microfilm sources. The film scanner was a modification of a microfiche scanner that also could accommodate aperture cards, 16 mm and 35 mm microfilm.

The scanned documents were stored on 12 inch optical disk media mounted in a jukebox. The jukebox configuration included two separate disk drives, and the jukebox was supported with a separate microprocessor server.

Several workstations were included in the ODISS configuration. A detailed view of the workstations is shown in Figure 7.4. Each workstation featured a microprocessor board and board level components supporting image compression, decompression, and high-resolution image display functions. The workstations were used for various purposes within the system. The scanner control

as installed at NARA. Overall con- within the system was performed by ssor on which a commercially avail- lent. The bus architecture was based is between the system manager and is design was selected to minimize r slower commercially standardized data transfer.

ustom software supported several dif- ves of the ODISS pilot system was available scanning equipment to han- dent sizes and paper weight, and docu- parate scanners were installed within eed paper scanner was used for pro- ty of document material. Clear poly- more fragile documents to prevent ons, which worked satisfactorily with nner operated at 200 dpi, scanning and ble-sided. Custom software supported indexing of individual documents.

Also provided. A flat bed paper scanner material that was too fragile to be . A third paper scanner, also a flat bed er pixel) and was provided to support ating various image processing and essing subsystem was used to exper- ment techniques to poor quality docu- orithms were applied to scanned 8 / of scanned material whose original

a the ODISS to evaluate the quality of y of microfilm sources. The film scan- scanner that also could accommodate ilm.

n 12 inch optical disk media mounted ncluded two separate disk drives, and microprocessor server.

the ODISS configuration. A detailed gure 7.4. Each workstation featured a mponents supporting image compres- image display functions. The worksta- thin the system. The scanner control

## OPTICAL DIGITAL IMAGE STORAGE SYSTEM (ODISS)

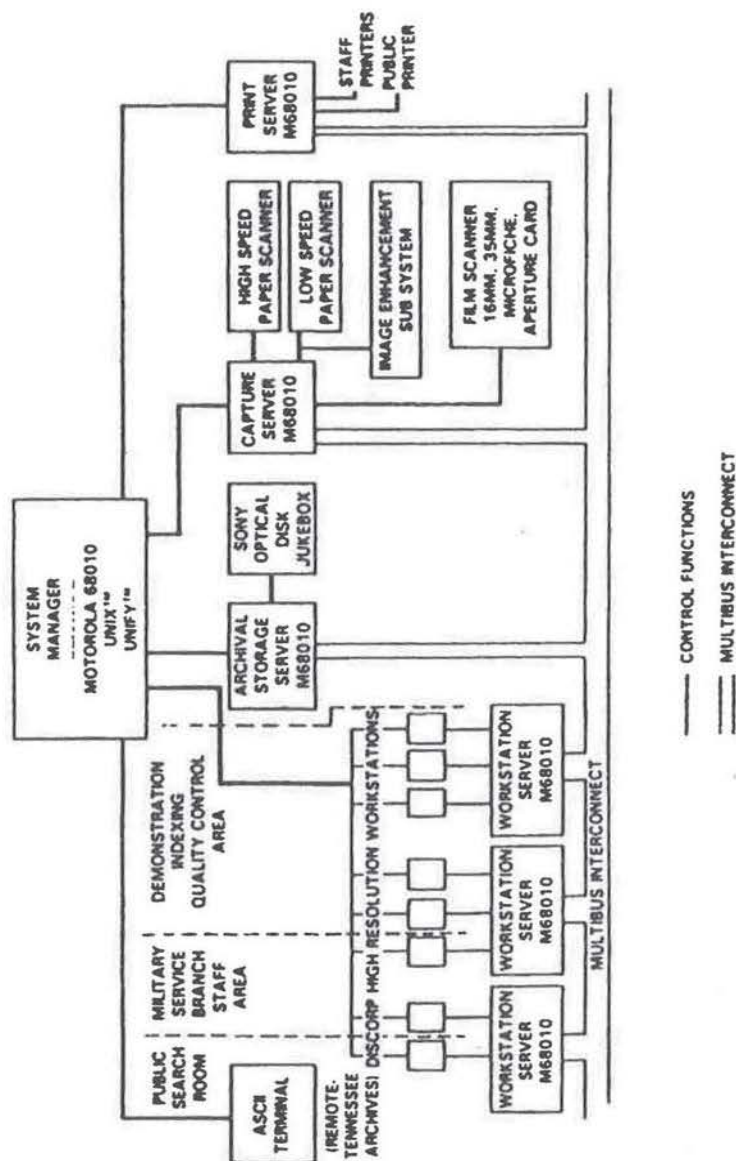
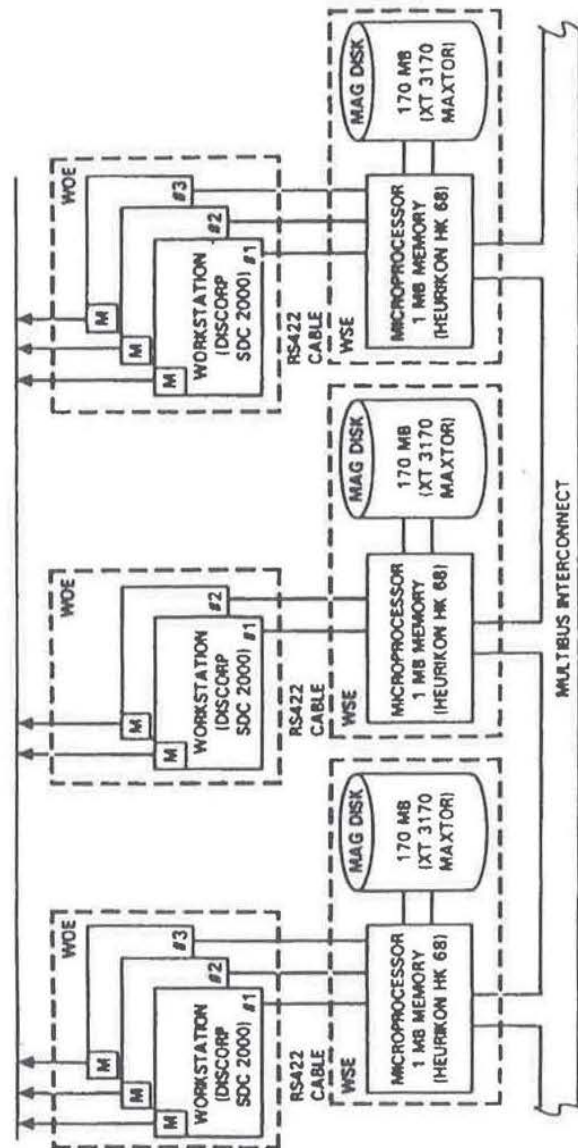


Figure 7.3 The NARA ODISS System. From Reference 1.



## TO SYSTEM MANAGER SUBSYSTEM



WSE - WORKSTATION ELEMENT  
 WSE - WORKSTATION SERVER ELEMENT

Figure 7.4 Details of the workstation configurations within ODISS. From Reference 1.

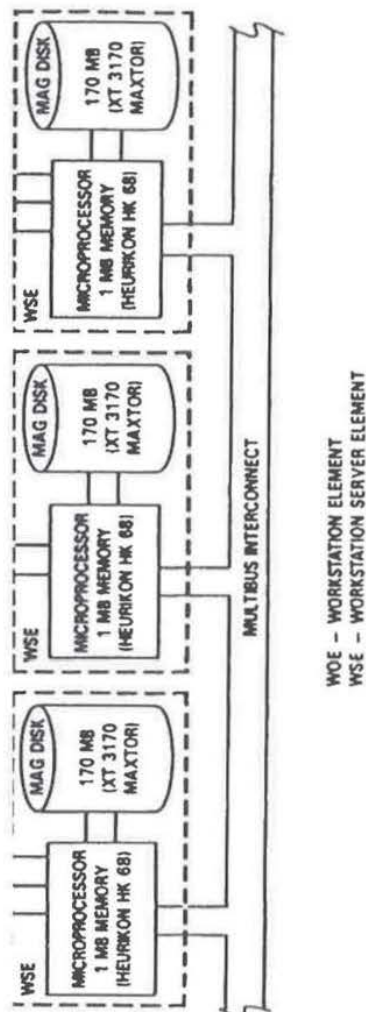


Figure 7.4 Details of the workstation configurations within ODISS. From Reference 1.

workstation incorporated software controlling the various scanners in the scanning subsystem. A quality control workstation was used to review scanned document images and the associated index data that had been entered during the scanning operation. The quality control operator would compare the scanned document image and associated index information on the same display screen (see Figure 7.1 for an example) with the original paper document and determine that the index information was correct and that the quality of the scanned image was acceptable. If the scanned material was acceptable, the scanned document would be entered into the archival database on optical disk; if not, the document would be returned for rescanning.

Workstations supporting the NARA staff in query and retrieval also incorporated custom software supporting those functions. Figure 7.5 shows the top-level menu presented at initiation of a database search by NARA staff members. In the lower left portion of the screen menu, staff members would enter as many fields as would be available to initiate a search. The system then utilized a commercially available database management system to search the index fields associated with the scanned document database. It would then present the results in the format shown in Figure 7.6. At that point, researchers could request image display of specific documents in the formats shown earlier in this section. The system also

Staff Workstation		National Archives Optical Digital Image Storage System Staff Workstation
Key	Action caused by the key	
F1	HELP	This Display allows you to search the ODISS system for a specific file to be viewed or printed.
F2	Begin Search of the Data Base	Any known values should be entered into the search area on the bottom left of this display. Fill in as much information as possible.
F3	Display Code Tables in Numeric Order	Code values available for all fields using numeric codes are displayed by positioning the cursor to the desired field, and then pressing F3. This displays the code table associated with that field listing both the numeric codes and their complete descriptions. Entering the code value, or pressing any function key will remove the display of the code tables, and allow you to continue filling out the search screen.
F4	Display Code Tables in Alpha Order	
F7	Clear all Fields on the Screen Below	
F8	Print Options	If the print options key allows you to print a file, or group of files whose File Control Number is already known. The file is directly sent to the printer, and is not displayed. This option also allows the printing of the code tables.
F9	Print the Entire Screen Exactly as it Appears	When you have finished entry of the search values for the data base search, you must press F2 to begin the search of the data base. A new screen will then show you those items in the data base that match the entered selection values.
F10	LOGOFF - Exit from the ODISS System	Pressing F10 will exit from the ODISS system. This will allow another user to enter their Employee Identification Number, and access the staff workstation.

TCN	
War	01 Civil War
State	01 Tennessee
Service	01 Confederate Army
Status	01 Active
Last Name	
First Name	
Middle Name	
Back In	
Back Out	
Payment	
Company (1)	
Company (2)	
Company (3)	
Company	

Figure 7.5 Top level menu used to initiate document database searches by NARA staff members using ODISS. From Reference 1.



P	P	FCN	Mar	St	St	Name	Rank	Reg	Co1	Co2	Co3	National Archives Optical Digital Image Storage System	
00000000	01	TH	01	01	01	BOND, JAMES M	001	001	001	001	001	Use the cursor keys to move the highlight bar to the file you wish to select from the list at left.	
00000000	01	TH	01	01	01	BOND, J W	001	001	001	001	001	A detailed description of the selected record will appear on the lower left screen. The CO# fields will be displayed with their text descriptions.	
00000000	01	TH	01	01	01	BOND, J C M	001	001	001	001	001	You may print or view any file that has been selected as described above. At any time pressing F10 will return to the initial search menu to allow entry of new search values.	
00000000	01	TH	01	01	01	BOND, JOHN M	001	001	001	001	001	Files that have been viewed or selected are marked with a P or V to the left of the FCN in the upper window.	
00000000	01	TH	01	01	01	BOND, JOHN M	001	001	001	001	001		
FCN 00000000: Total number of pages 0001												FINISH - Found 0 matches!	
You are viewing data base match 1												F1 - Help	
Last Name: BOND												F2 - Display selected file	
First Name: JAMES												F3 - Print Options	
Middle Name: M												F4 - Print the entire screen exactly as it appears	
Rank In: 001 Lieutenant												F10 - CANCEL - returns to the search menu to enter new search values	
Rank Out: 001 Lieutenant													
Regiment: 011 Grenade (Buchworth's) Company													
Company (1): 001 D													
Company (2): 001													
Company (3): 001													
Branch: 001													

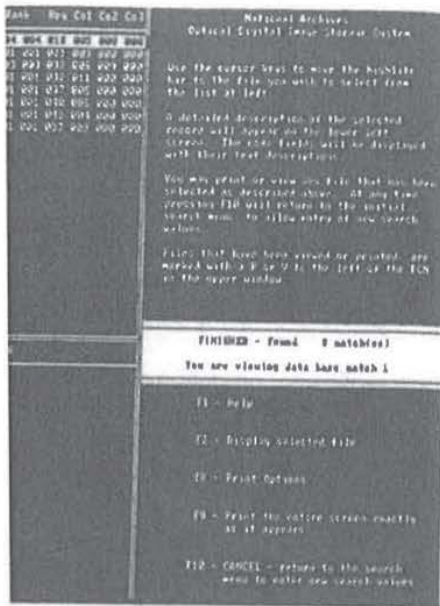
**Figure 7.6** Results of database query initiated as shown in Figure 7.5. The staff member can now display selected documents by using functions keys as shown in the lower right-hand corner of the display. From Reference 1.

supported remote access to index data (but not image access) at the Tennessee state archives, and it provided image display at the public search room at NARA.

## Summary

ODISS was used to scan the Tennessee CMSR database and was then used to support staff query and retrieval functions within a controlled and monitored environment. The final ODISS report contains many significant conclusions and cost comparisons between the use of electronic document management technology and the alternative use of paper files or microfilm. All of the results of this pilot test are too voluminous to be reprinted here, but following are several of the conclusions summarized in Reference 1:

- The ODISS Project demonstrated a three-fold reduction in the amount of staff time required to search and retrieve name index files of the Tennessee Confederate CMSR...
- The ODISS search and retrieval system provides a greatly enhanced access capability to the Tennessee Confederate CMSR. Extension of this search and retrieval capability to other CMSR, Pension, and Bounty Land



initiated as shown in Figure 7.5. The staff member by using functions keys as shown in the lower reference 1.

ta (but not image access) at the Tennessee display at the public search room at NARA.

see CMSR database and was then used to actions within a controlled and monitored contains many significant conclusions and electronic document management technologies or microfilm. All of the results of this printed here, but following are several of nce 1:

ed a three-fold reduction in the amount of 1 and retrieve name index files of the

val system provides a greatly enhanced see Confederate CMSR. Extension of this other CMSR, Pension, and Bounty Land

records is likely to lead to greater use of the records as new questions of historical research are formulated...

- From the cost analysis...it appears that any conversion of paper records to an alternate form cannot currently be justified purely on the basis of cost alone. However, the ODISS project also identified other intangible benefits, such as improved image legibility, improved timeliness and accuracy of access, an enhanced retrieval capability, reduction of storage space requirements, and reduced or eliminated handling of original documents. A conversion of records to an alternate form may be justifiable on a basis other than reduction of costs. Each case must be decided on its own individual merits...

The ODISS project also demonstrated the effectiveness of implementing a pilot system with well defined objectives addressing a subset of a total operation for use in evaluating the utility and potential cost benefits of this new technology prior to commitment to the new technology on a much broader scale within an existing operation.

## ENGINEERING DATA MANAGEMENT SYSTEMS

Digital document management systems capable of supporting engineering drawing databases of hundreds of thousands of engineering drawings are emerging as a special class of general electronic document management systems. This section describes two different engineering document management systems produced by different vendors for customers with specific requirements. In both cases described here, it was necessary to perform backfile capture of an existing data base of engineering documentation; both systems include interfaces with Computer Aided Design (CAD) workstations for capture of current design information that is generated in computer compatible format.

### Engineering Drawing Repository

The Integrated Automation Division of Litton Industrial Automation Systems, Inc., has implemented an Engineering Data Repository system for a worldwide heavy equipment manufacturer. The system is shown in Figure 7.7. This system is typical of current open architecture standards-compliant designs. A variety of equipment from several manufacturers has been integrated into a unified system, with communications paths and interfaces that exploit existing standards.

The system supports electronic capture, storage, revision management, archiving, distribution, viewing, and communication of engineering drawings, designs, and change documentation. The database managed by the system consists of both scanned documents and data files entered directly from Computer Aided Design (CAD) systems. The scanned document material was scanned



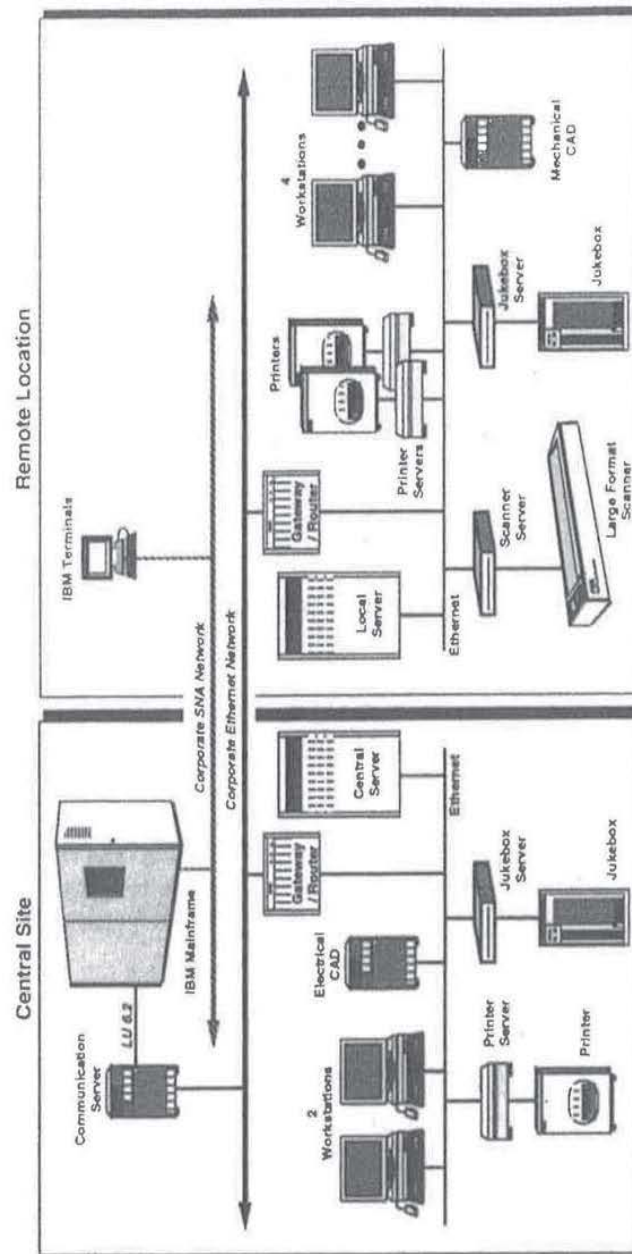


Figure 7.7 Block diagram of the Engineering Drawing Repository System, showing integration of scanners, optical disk storage device, Computer Aided Design workstations, document workstations, and central mainframe used to support over all document management and routing within the system. Courtesy of Litton Integrated Automation Division.

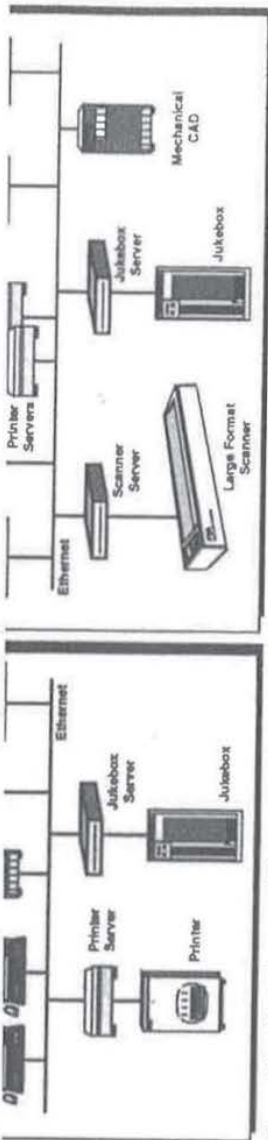


Figure 7.7 Block diagram of the Engineering Drawing Repository System, showing integration of scanners, optical disk storage device, Computer Aided Design workstations, document workstations, and central mainframe used to support over all document management and routing within the system. Courtesy of Litton Integrated Automation Division.

from engineering drawings on paper and on aperture cards. Scanning from paper required use of a large format scanner designed to scan engineering size drawings, as shown in the figure.

The central control of data flow within the system is performed by the mainframe computer. The mainframe computer also supports database management functions, including user query. Users query the mainframe computer to locate a desired drawing or engineering change notice. A function key control at the user workstation initiates display of the document on the workstation display. The majority of engineering drawings are too large to be viewed at full resolution on even a high resolution display. For that reason, pan and zoom functions are provided to the users at the workstations.

The following functions are performed by this system, supported by applications software:

- View and print documentation
- Archive and backup storage
- CAD conversion
- Full size drawing scanning
- Input from a scanning service bureau
- Control of pending changes to drawing
- Electronic mark-up of drawings and document
- Manufacturing planning
- Electronic distribution
- Engineering document preparation
- Control and release of design drawings

The following objects are (or soon will be) managed by the system shown in Figure 7.7:

- Requirements specifications
- Product drawings
- Change markups
- CAD files and libraries
- Engineering analysis and support data
- Test specifications
- Simulation models
- Technical publications
- Purchase specifications



This system is typical of an engineering drawing management application involving management of older drawings that reside on paper or aperture cards and newer drawings that have been produced by electronic CAD systems. Backfile capture is required to convert the older drawing database to digital image format. An interface with the newer CAD technology is required to support current ongoing design and manufacturing activities. This system will also support workflow requirements within the organization for which the system was developed.

### **Syncrude Canada Ltd., Computer-Aided Engineering and Design (CAED) System**

Syncrude Canada Ltd. is the world's largest producer of synthetic crude oil. The company operates near Fort McMurray, Alberta, mining the Athabasca Oil Sands. This oil sand deposit contains an estimated 1.3 trillion barrels of bitumen. Syncrude mines the oil sand from an open pit. The company extracts bitumen, upgrades it to synthetic crude oil, and transfers the crude oil to refineries by pipeline. More than 25 miles of conveyer belts are required to transport the sand from the pits to the processing centers. This operation provides Canada with 11 percent of its daily crude oil requirements.

Figure 7.8 shows an example of an engineering drawing, and Figure 7.9 shows the aperture card retrieval system in use at Syncrude prior to implementation of an electronic document management system. The decision to move to an electronic document management system was based on a need to improve the productivity of engineering operations at Syncrude. A survey of over 300 engineers, design drafters, maintenance planners, and inspectors at Syncrude indicated that between 15 and 25 percent of their time was spent identifying and



**Figure 7.8** Typical Syncrude engineering drawing. Courtesy of Integraph Corporation.

rawing management application reside on paper or aperture cards used by electronic CAD systems. Older drawing database to digital D technology is required to support activities. This system will also organization for which the system

## Aided Engineering and

ducer of synthetic crude oil. The Alberta, mining the Athabasca Oil field 1.3 trillion barrels of bitumen. The company extracts bitumen, sends the crude oil to refineries by rail. It is required to transport the sand. Extraction provides Canada with 11

engineering drawing, and Figure 7.9 at Syncrude prior to implementation. The decision to move to an electronic system was based on a need to improve the productivity. A survey of over 300 engineering inspectors at Syncrude indicated that more time was spent identifying and



Figure 7.9 courtesy of Integraph Corporation.

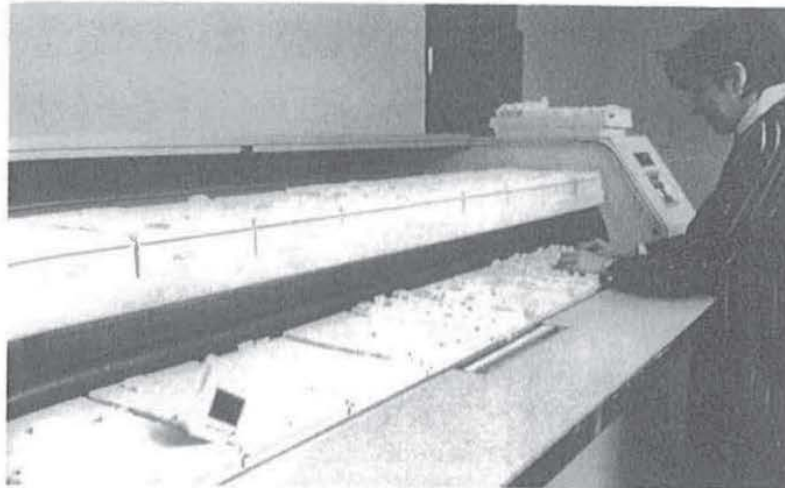


Figure 7.9 Engineering drawing aperture card filing system in use at Syncrude prior to implementation of the CAED system. Courtesy of Integraph Corporation.

retrieving engineering drawings and aperture cards. This survey was a contributing factor in making the decision to move to an electronic document management system. Integraph Corporation was selected to implement the imaging and CAD portion of the CAED system for Syncrude.

The project has been implemented in two phases. The first phase involved delivery and operation of the scanning subsystem used to capture the existing backfile of aperture cards and engineering drawings. A customized user interface was developed to support the scanning function. A major requirement was development of software to utilize Syncrude's existing drawing numbering system database, and creation of a method to link images to database records. Production scanning (primarily of aperture cards) was initiated in October 1990, and as of May 1992, more than 225,000 drawings had been converted. After completion of backfile capture, approximately 450,000 engineering drawings will be converted to digital image format.

Figure 7.10 shows a block diagram of the CAED system. Figure 7.11 shows one of the retrieval and display workstations within the system. In order to implement the total system, a significant amount of existing Syncrude equipment had to be integrated with the new electronic document management capabilities. The completed system incorporates over 250 personal computers, over 50 Integraph CAD workstations, and 11 IBM RISC 6000 engineering workstations that were in place prior to the transition to electronic document management technology. Integraph added the scanning subsystem (aperture card and large format paper scanning), document storage capability, and customized data management systems, and user interface software.



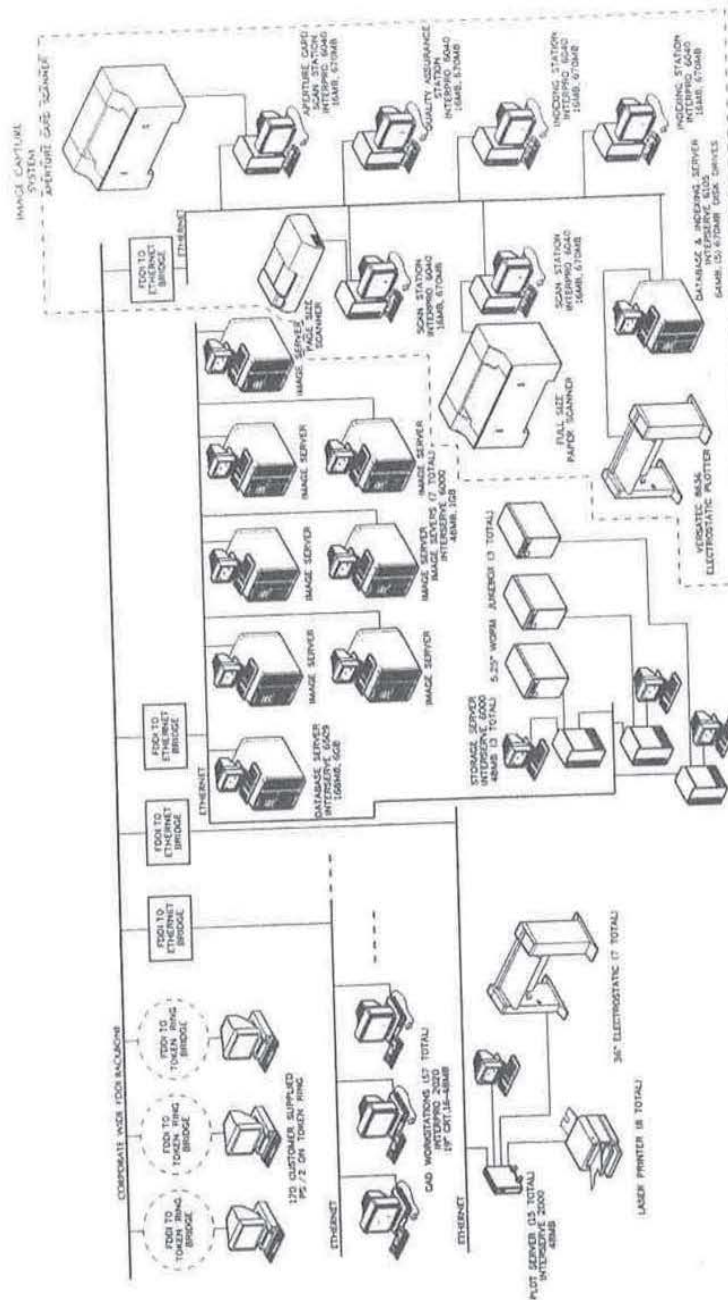


Figure 7.10 Block diagram of the Syncrude CAED system. Integration of the electronic document management capability with existing Syncrude equipment and existing engineering drawing database management systems was required. Courtesy of Integraph Corporation.

Figure 7.11 L  
Courtesy of Int

The system  
ing number  
approval of  
new and ex  
CAD workst  
neering infor  
material conv  
data producer

The use of e  
technology e  
applications  
strated in the  
tions continu  
technology h  
reducing the

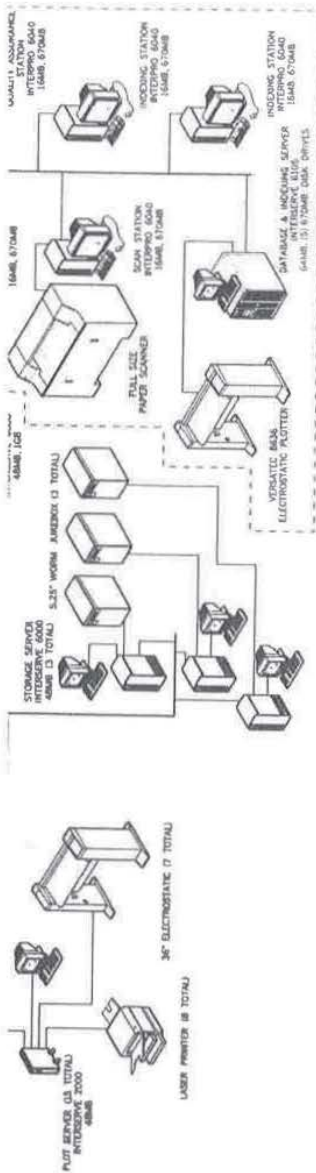


Figure 7.10 Block diagram of the Syncrude CAED system. Integration of the electronic document management capability with existing Syncrude equipment and existing engineering drawing database management systems was required. Courtesy of Integraph Corporation.



Figure 7.11 User query, retrieval, and display workstation within the CAED system. Courtesy of Integraph Corporation.

The system had to be designed to use Syncrude's existing engineering drawing numbering systems. It had to support functions such as markup, review, and approval of engineering documentation as well as user interface support at both new and existing workstations. An interface with engineering data created via CAD workstations was provided. This resulted in the ability to retrieve engineering information from a total database, which includes the scanned backfile material converted from paper or aperture card media and the more current CAD data produced on digital workstations.

## BANKING SYSTEMS

The use of electronic document management technology is increasing as the technology evolves and individual institutions implement electronic systems in applications where the cost benefit or productivity improvement can be demonstrated in their operations. The volume of paper processed by banking institutions continues to increase significantly, and electronic document management technology has become a viable technology for improving productivity and reducing the costs associated with this processing.



There are several banking applications in which electronic document technology is being applied. These include:

**Check Processing.** Banking institutions have traditionally employed high speed equipment for sorting and processing large volumes of paper checks on a daily basis. These systems have included operator stations where check amounts are encoded and other equipment involving high speed paper transport and interaction with operator functions. Electronic document management systems are now available that support several different functions associated with bank check processing that support processing of the paper sizes associated with this application.

**Credit Card Operations.** Credit card operations can involve large volumes of remittance forms that must be processed on a daily basis. Although electronic point-of-sale data entry has been used to address some aspects of credit card activity, most banks must still process paper documents on a daily basis.

**Loan Applications.** There are large volumes of paper documentation associated with loan applications and loan processing. The documents involve both fixed format forms and associated documentation that may not be in a predetermined format. It is often possible to utilize more electronic document management systems in these applications that are not as highly specialized as systems developed to support check processing or credit card operations.

**Signature Verification.** Systems are available that retrieve and display scanned signature cards as part of the process of operator verification of customer signatures.

**General correspondence and Customer Service Applications.** For these applications, it is often cost effective to implement a general purpose document processing system that may be customized to the specific work flow within an individual institution. These systems are used to support electronic storage and display of general office correspondence and other document material.

This section will describe check processing applications in more detail and provide one example of an operational system that has improved productivity in one specific banking application.

## Check Processing

For many years, banks have been using a variety of equipment in check processing that is not based on image processing technology. Recently, commercially available systems incorporating electronic document management technology have emerged that support specific functions involved in daily processing of checks.

Conventional check processing involves several steps within a typical banking operation. All checks received at a branch are encoded by the bank utilizing

tions in which electronic document technol-

stitutions have traditionally employed high processing large volumes of paper checks on a dedicated operator stations where check amounts involving high speed paper transport and Electronic document management systems perform different functions associated with bank processing of the paper sizes associated with this

card operations can involve large volumes processed on a daily basis. Although electronic processing to address some aspects of credit card processing on a daily basis.

large volumes of paper documentation and loan processing. The documents involve electronic documentation that may not be in a pre-able to utilize more electronic document processing that are not as highly specialized as processing or credit card operations.

is are available that retrieve and display the process of operator verification of

**Customer Service Applications.** For these to implement a general purpose document processing system customized to the specific work flow within an are used to support electronic storage and retrieval and other document material. processing applications in more detail and a system that has improved productivity in

g a variety of equipment in check processing technology. Recently, commercially available electronic document management technology functions involved in daily processing of

olves several steps within a typical bank-branch are encoded by the bank utilizing



Figure 7.12 Illustration of the addition of the check amount to the MICR strip during the encoding process. Top, check as written on pre-printed check form. Bottom, check after encoding process has added the amount to the MICR strip. Checks compliments of Unisys Corporation.

equipment that allows an operator to read and key-enter the amount. The amount is then added to the Magnetic Ink Character Recognition (MICR) strip at the bottom of the check. Deposit balancing and verification is also generally performed at this step. An example of the encoding process is shown in Figure 7.12. At the top of Figure 7.12 is the check as written by the individual. The pre-printed checks used by consumers are provided with a MICR imprint that contains the bank routing number, the account number, and the check number (1551

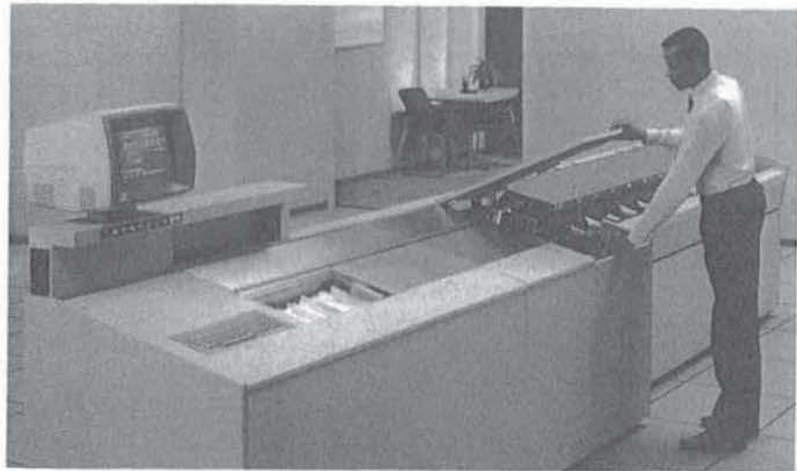


Figure 7.13 High speed check reader/sorter unit. Courtesy of Unisys Corporation.



in this example). The bottom of Figure 7.12 shows the same check after the encoding process. The check amount (\$125.00 in this example) has been added to the MICR strip through the encoding process.

A high speed reader/sorter is generally utilized at this point in the processing. This device reads the MICR line that is imprinted on every check and sorts the check into one of several output pockets based on the sort criteria of the bank. Multiple sorts are often required to support different requirements. A final sort on the account number is used to sort cancelled checks for transfer back to the customer. A microfilming camera is generally an option on most reader/sorters, providing the capability of filming checks during the sorting process for archival storage purposes. One example of a high speed reader/sorter is shown in Figure 7.13.

Several correction steps are required in conventional check processing systems. Checks that exhibit problems reading the MICR line must be reentered. Balances must be reconciled, involving a process of isolating individual problems (damaged checks, miscoded checks, checks stuck together, etc.), generally performed by hand.

Image-based check processing systems support the more labor intensive operations in a typical check processing system. Account entry workstations are used to enter amounts from scanned, digitized imagery displayed on the workstation display. Images are typically scanned, compressed, and stored on magnetic disk, then retrieved for amount entry and MICR encoded after data entry and balancing. Application software can be used for deposit reconciliation at the data entry point. Some commercially available systems now provide optical character recognition capability that reads the "courtesy amount" from handwritten entries on the original checks. The legal amount for which a check is drawn is based on the handwritten value entered on the check. The courtesy amount is the numerical amount for which the check is drawn that is written on the "pay to the order of" line. Accuracy of these devices can be expected to increase in the future as check form designs improve and recognition technology improves. This feature is often referred to as *Courtesy Amount Recognition* (CAR).

High speed reader/sorter/endorsers are used to physically process and sort checks. Image workstations are used to retrieve digitized check images to resolve problems. Workstations incorporate image processing and manipulation functions, such as zoom, pan, rotation, and limited image enhancement to support operator analysis of hard to read information from the original check. Workstations can also be used to fill in data missing or unreadable on the MICR line. Some systems produce an image check statement by performing a digital (rather than physical) sort and generating an image of all checks processed for a given account.

One example of a currently operational check processing system is a system installed at Signet Bank of Virginia by Unisys Corporation. This system is being implemented in a phased manner, involving cost benefit analysis performed after each phase of the implementation. The first phase involved evaluation of data entry performance of image systems. Signet operators were trained on

2 shows the same check after the 00 in this example) has been added asss.

lized at this point in the processing. rinted on every check and sorts the sed on the sort criteria of the bank. different requirements. A final sort lled checks for transfer back to the an option on most reader/sorters, pro- the sorting process for archival stor- der/sorter is shown in Figure 7.13. onventional check processing sys- the MICR line must be reentered. rocess of isolating individual prob- ecks stuck together, etc.), generally

pport the more labor intensive oper- Account entry workstations are used imagery displayed on the workstation ressed, and stored on magnetic disk, ncoded after data entry and balanc- posit reconciliation at the data entry ms now provide optical character y amount" from handwritten entries which a check is drawn is based on The courtesy amount is the numeri- at is written on the "pay to the order expected to increase in the future as n technology improves. This feature gnition (CAR). used to physically process and sort retrieve digitized check images to image processing and manipulation limited image enhancement to sup- formation from the original check. missing or unreadable on the MICR k statement by performing a digital i image of all checks processed for a

check processing system is a system ys Corporation. This system is being ng cost benefit analysis performed ie first phase involved evaluation of i. Signet operators were trained on

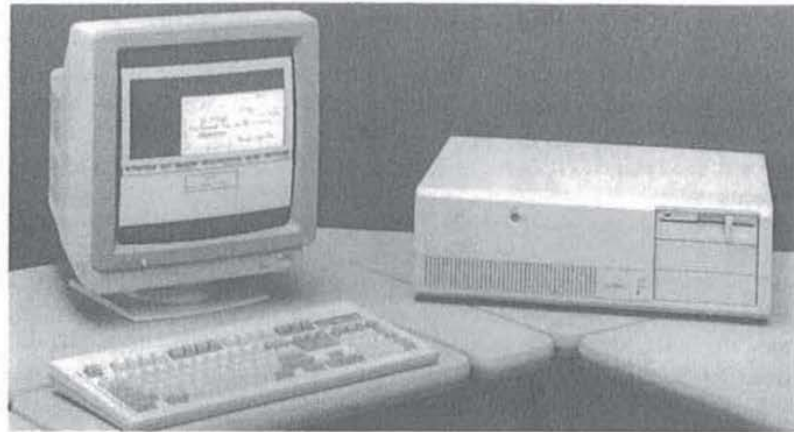


Figure 7.14 Digitized check image decompressed and displayed for amount entry by operator. Courtesy of Unisys Corporation.

account entry workstations, as shown in Figure 7.14. After ten six-hour sessions, six of Signet Bank's proof-of-deposit operators averaged 2,515 documents per hour with an average error rate of 1.15%. This compares with conventional check data entry systems where operators averaged 1,200 documents per hour with an average error rate of 1%.

The system also includes high speed reader/sorter units, and operator image processing workstations, as shown in Figure 7.15. When fully configured, the final system should be capable of processing approximately 700,000 checks daily.



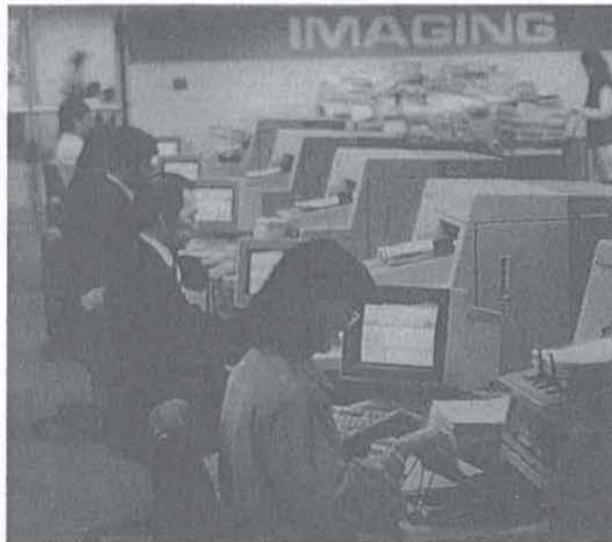
Figure 7.15 Image workstations used in the Signet Bank system. Courtesy of Unisys Corporation.



## MEDICAL CLAIMS PROCESSING

Medical claims processing operations often involve centralized sites that process thousands of incoming documents daily. The incoming documentation includes standardized forms, receipts, and associated documentation. The claims processing operation involves operator evaluation of the claim, generation of correspondence to the individual filing the claim, determination of payment, generation of checks, and production of payment records and accounting information. Insurance companies generally use measurements of the time between receipt of a claim and delivery of payment to the customer as a basis for evaluating their performance and as a basis for comparison against competing insurance companies. The need for rapid processing of large volumes of information, and the requirement for long term retention of records of the information processed, make medical claims processing a natural candidate for productivity improvement through the use of electronic image technology.

One example of the use of electronic imaging to improve claims processing performance is the system installed at Empire Blue Cross and Blue Shield of New York by Sigma Imaging Systems Inc. Empire decided to implement imaging within their customer service and claims processing departments. As the nation's largest private not-for-profit insurer, the company receives 125,000 paper claims each day across five sites in New York State. The use of imaging has dramatically reduced the processing cycle of the average claim.



**Figure 7.16** High speed claims scanning operation at Empire Blue Cross and Blue Shield. Figure Courtesy of Sigma Imaging Systems, Inc.

## PROCESSING

involve centralized sites that process incoming documentation includes documentation. The claims processing of the claim, generation of correspondence, determination of payment, generation of orders and accounting information. Measurements of the time between receipt of documents as a basis for evaluating their accuracy against competing insurance companies, volumes of information, and the records of the information processed, candidate for productivity improvement technology.

Imaging to improve claims processing at Empire Blue Cross and Blue Shield of New York State. The company decided to implement imaging in its claims processing departments. As the insurer, the company receives 125,000 claims per year from New York State. The use of imaging reduced the average claim.



ation at Empire Blue Cross and Blue Shield of New York State. Figure courtesy of Sigma Imaging Systems, Inc.



Figure 7.17 View of one area of the claims processing system installed at Empire Blue Cross and Blue Shield. Figure courtesy of Sigma Imaging Systems, Inc.

OmniDesk software developed by Sigma Imaging Systems Inc. is used to control 26 high speed scanners to scan, automatically index, and route an average of 2,000 images per hour from each scanner. Empire currently has over 1,600 image-enabled workstations in operation processing over a quarter of a million imaged pages each day (each claim received can consist of more than one page of material). A portion of the high speed scanning operation is shown in Figure 7.16.

Claims processing for one of Empire's operations consisted of fifteen separate worksteps. At each step, batches of paper were controlled, distributed, and then delivered to the next step in the flow. Information needed to be analyzed, coded, and at the final step, keyed into a mainframe-based claims processing system. When the electronic imaging system was installed, the number of worksteps was reduced to one, and claim suspension rates were cut by one-third. At one site, Empire eliminated 70 to 80 percent of the claims examiner workload by using Sigma software to reengineer and automate claims processing workflows. In a three year period, Empire installed over 1,600 image-enabled workstations to support this application. Figure 7.17 shows one processing area at Empire.

## LARGE VOLUME STORAGE AND RETRIEVAL SYSTEM—THE AUTOMATED PATENT SYSTEM

The Automated Patent System (APS) installed at the Patent and Trademark Office, United States Department of Commerce, is an extensive distributed system that includes image workstations, mainframe and personal computers, large image and text databases, text search, image storage and retrieval, file servers,



wide-area and local area networks, and network and system management subsystems. System design and development started with the award of a design, development, implementation and operations contract to PRC, Inc., in 1984.

At this writing, the system provides image storage and retrieval services for 75 workstations and 300 text search terminals. In addition, 70 Macintosh computers and 150 PC's are connected to the system for both text search and office automation. The system supports high speed text search on all post-1971 U. S. patents, abstracts of Japanese and European patents issued after 1980, and patent related technical literature. At this writing, users can retrieve images of more than 750,000 U. S. patents. When PRC completes the image load process, all 6 million U. S. patents will be available for retrieval by the image workstations. Later, 9 million European and Japanese patents will also be available for recall in image form. At completion, the APS will constitute the world's largest image database.

A system diagram of the APS is shown in Figure 7.18. The APS design is based on a distributed architecture with patent text, text search software, and image index files maintained on a large mainframe computer. Images and related files are stored on multiple, distributed optical and magnetic disk based file servers. These devices and associated text search and image workstations are interconnected using a local area network with gateways providing access to external networks. A three-level, hierarchical local area network ("PTOnet") is being implemented by PRC to provide expansion capability to support a total bandwidth of approximately 1 Gbit/sec.

Current APS plans include the installation of approximately 2,000 image workstations and additional image file servers. PTOnet will ultimately support approximately 5,000 devices.

Document images are stored on WORM media in CCITT Group-IV compressed form, requiring approximately 85 KBytes per image. When completed, the image database will hold approximately 200 million images. Image data is stored at two different resolutions, 150 dpi and 300 dpi. The 150 dpi image database is stored on rapid access devices (RAD's) that incorporate four optical disk drives, providing rapid access to four disks per RAD. The 150 dpi image data is used for display at the high resolution image workstations that display the imagery at the 150 dpi resolution. The 300 dpi image database is stored on high density devices (HDD's) that consist of optical disk jukeboxes holding 50 WORM disks each. The HDD's support printing requirements at 300 dpi resolution, support 2:1 image scaling at the image workstations, and serve as a backup to the 150 dpi database stored on RAD's.

The system was designed to support three critical performance requirements: (i) search of the entire non-image (text) database in a maximum of 30 seconds, (ii) display of the first page of image data in a requested patent subclass in less than 10 seconds, and (iii) display of subsequent pages at a rate of one page per second.

work and system management submitted with the award of a design, contract to PRC, Inc., in 1984. storage and retrieval services for . In addition, 70 Macintosh computers for both text search and office text search on all post-1971 U. S. patents issued after 1980, and patent users can retrieve images of more patents the image load process, all retrieval by the image workstations. This will also be available for recall to constitute the world's largest image

Figure 7.18. The APS design is a text, text search software, and mainframe computer. Images and optical and magnetic disk based search and image workstations are with gateways providing access to local area network ("PTOnet") is a system capability to support a total

of approximately 2,000 image units. PTOnet will ultimately support

media in CCITT Group-IV computers per image. When completed, 100 million images. Image data is at 300 dpi. The 150 dpi image data that incorporate four optical disk drives. The 150 dpi image data is at image workstations that display the image database is stored on high optical disk jukeboxes holding 500 gig requirements at 300 dpi resolution workstations, and serve as a backup

critical performance requirements: time in a maximum of 30 seconds, to requested patent subclass in less than 10 pages at a rate of one page per

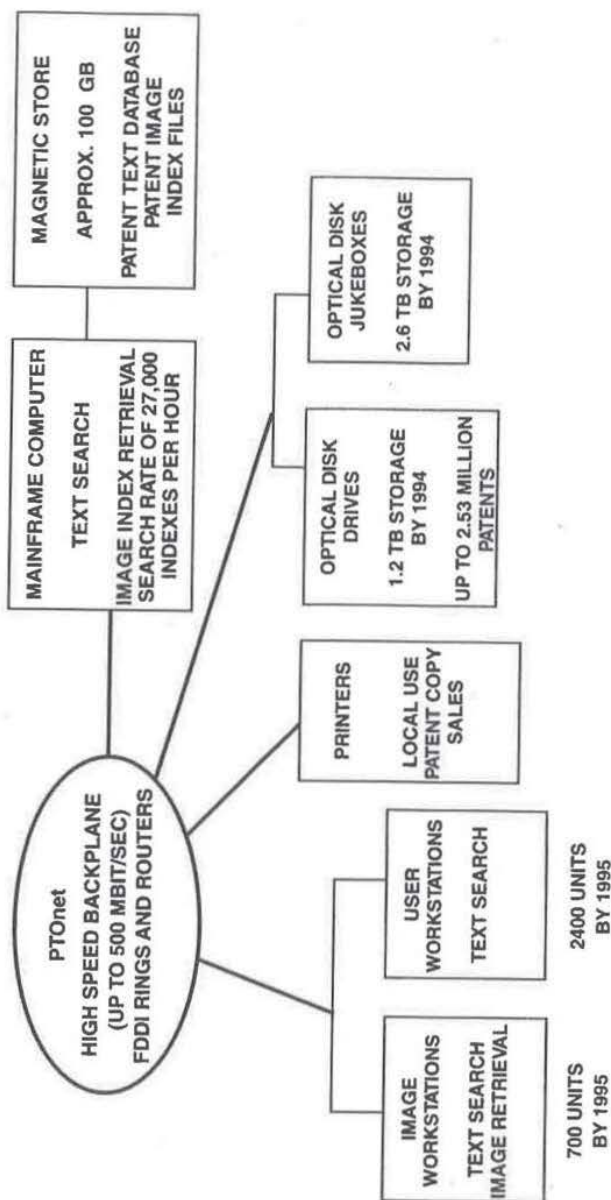


Figure 7.18 Automated Patent System at the U. S. Patent and Trademark Office. Figure courtesy of PRC.



The APS was one of the first major electronic imaging procurements released by the Federal Government. PRC has incorporated emerging technology into the system as the industry has evolved, and it has been supported by over 25 hardware vendors and 17 software vendors in building this system.

### References

1. National Archives and Records Administration Archival Research and Evaluation Staff, *Optical Digital Image Storage System Project Report*, 1991 (available from National Technical Information Service, Springfield, VA).



## CHAPTER 12

# Bar Code Technology

Bar code technology has become widely used in many applications as a method of entering a limited amount of alphanumeric data in a high reliable and error-free manner. Bar codes can be interpreted by hardware or software systems, and the information encoded using bar codes is almost guaranteed to be error free after hardware or software decoding of the bar code pattern. Bar codes are used widely in inventory control applications ranging from supermarkets to industrial warehouses. The most familiar use of bar code technology is in retail sales (including supermarkets) where bar code scanners utilized at the point of sale correctly interpret the product or item being sold and provide a link to computer based data management systems which provide current item pricing and instantaneous updating of inventory control files.

This chapter describes the general methods used for encoding alphanumeric information using bar code technology, and it summarizes examples of the use of bar code technology in electronic document management systems. The description of basic bar code techniques illustrates the features of bar coding that lead to high reliability and almost zero error rate in interpreting bar coded data. The two basic types of bar coding, discrete and linear coding, are described.

There are many standards for encoding alphanumeric information using bar code technology. The most common method used in supermarkets is based on the Universal Product Code (UPC) standard developed in the 1970s. This code provides a method of encoding the specific manufacturer and product identifier for consumer products. This standard is not generally used in electronic document management systems. This chapter contains descriptions of the two most commonly utilized bar code standards in electronic document management sys-



tems, Code 39 and Code 128. Readers seeking more detailed information regarding bar code systems, standards, and specialized hardware are referred to more complete references.<sup>1</sup>

## BASIC BAR CODE CONCEPTS

The basic concept of bar code information encoding is shown in Figure 12.1. Information is encoded using a series of parallel black bars and white spaces. The information content and logical structure of the information is contained in the width of the bars and spaces. Bar codes are generally read in consumer applications using special scanners that incorporate lasers. The reflected light signal from the bar code pattern is interpreted and decoded within the scanner. One reason that this form of encoding is reliable is that the scan mechanism is not required to be perfect. The scan of a bar code can be performed at any angle, since the relative widths of the bars and spaces is preserved at any orientation relative to the scan direction. Additional encoding within a specific bar code standard provides parity checking, logical structuring within the encoded information to insure against misinterpretation, and other error elimination processes.

There are two basic types of character encoding used in bar codes, *discrete* and *continuous*. Discrete character bar code symbology encodes each character separately, with separation gaps between each successive character. Continuous encoding does not involve gaps between characters. When characters are encoded using continuous coding, every character begins with a bar and ends with a space. Figure 12.2 shows the difference between discrete and continuous coding.

There are several basic definitions used in describing the variety of bar code encoding schemes currently in use. A *bidirectional* code is one that can be scanned either from left to right or from right to left without affecting the data that is decoded. The majority of bar codes used in electronic document imaging

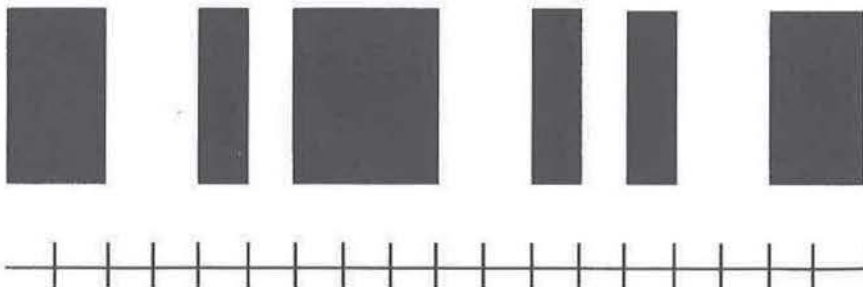


Figure 12.1 Basic bar code technique. Each black or white bar in this example is from one to three units thick. The information content is encoded within the sequence of widths of the alternating black and white bars.



Figure  
Contin

applic  
arbitra  
delimi  
variab  
same  
ing sc  
the an

(n,k)

A con  
are (n  
charac  
spaces  
(9,2) c  
indica  
that is  
limita  
and tw  
In t  
encod

rs seeking more detailed information and specialized hardware are referred to

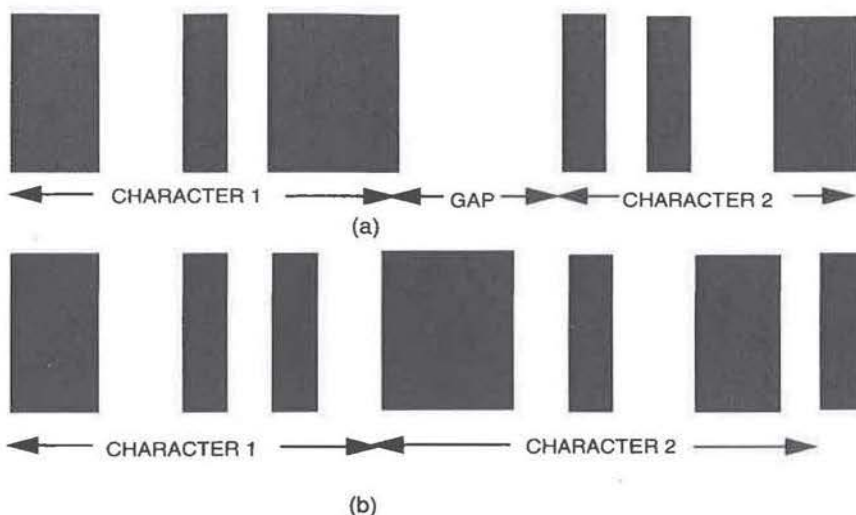
## DE CONCEPTS

tion encoding is shown in Figure 12.1. of parallel black bars and white spaces. icture of the information is contained in codes are generally read in consumer t incorporate lasers. The reflected light preted and decoded within the scanner. s reliable is that the scan mechanism is bar code can be performed at any angle, l spaces is preserved at any orientation al encoding within a specific bar code al structuring within the encoded infor- 1, and other error elimination processes. er encoding used in bar codes, *discrete* ode symbology encodes each character i each successive character. Continuous een characters. When characters are r character begins with a bar and ends erence between discrete and continuous

ed in describing the variety of bar code *bidirectional* code is one that can be ight to left without affecting the data s used in electronic document imaging



black or white bar in this example is from itent is encoded within the sequence of rs.



**Figure 12.2** (a) Discrete coding, where spaces separate each encoded character (b) Continuous coding, each character begins with a bar and ends with a space.

applications are *variable length* codes. This means that a string of characters of arbitrary length can be encoded, and embedded control characters are used as delimiters to indicate the end of a set of encoded characters. The opposite of a variable length encoding scheme is a *fixed length* encoding scheme, where the same number of characters is encoded for each bar code. A fixed length encoding scheme eliminates the need for delimiting characters, but it is less flexible in the amount of data that can be encoded within each bar code symbol.

## (n,k) Encoding Schemes

A commonly used set of bar codes in electronic document management systems are (n,k) codes. In an (n,k) code, n indicates the number of spaces within each character that can be either black or white and k indicates the number of bars (or spaces) within each character. As an example, a few characters encoded with a (9,2) continuous coding scheme are shown in Figure 12.3. A short computation indicates that a (9,2) encoding scheme can encode 56 discrete characters, since that is the maximum number of possible patterns that can be achieved using the limitations of the (9,2) encoding scheme—a limit of nine spaces, and two black and two white bars within the nine spatial modules.

In the general case, the total number of possible patterns in a particular (n,k) encoding scheme can be computed as:

$$\frac{(n-1)!}{(2k-1)! (n-2k)!}$$



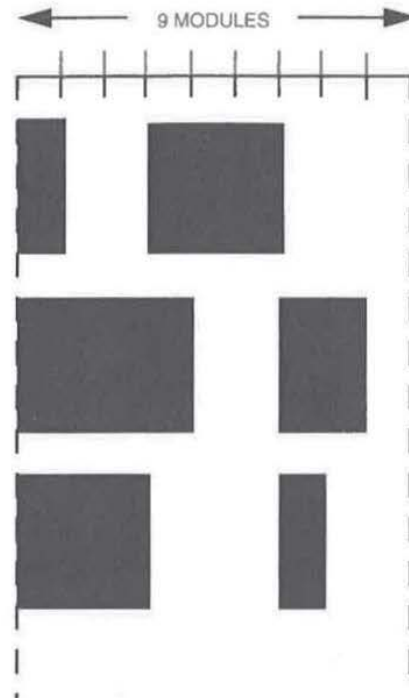


Figure 12.3 Examples of three different characters encoded using a (9,2) (n,k) encoding scheme.

There are two codes commonly used in electronic document management systems, Code 39 and Code 128. They are described in the next two sections.

### Code 39

Code 39 (sometimes called the "3 of 9" code) is a discrete variable length code that is widely used to encode alphanumeric data.<sup>2</sup> Every character encoded using Code 39 contains five bars and four spaces (the "9" in Code 39). Of the nine elements, three of the elements (bars or spaces) are wide (the "3" in Code 39) and the other six are narrow. The code definition provides for forty-four different alphanumeric characters. One of the forty-four characters (the asterisk) is reserved for use as a "start" and "stop" character, indicating the start and end of each encoded character string. The remaining forty-three characters consist of the 26 letters, 10 numerals, and the following symbols: - (period) (space) \$ / + and %. Figure 12.4 shows the Character "A" encoded using Code 39. Figure

Figure 12.5

shows the character "stop" encoded using Code 39, with

CH
1
2
3
4
5
6
7
8
9
0
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
]
{
}
~
*

Figure 12.6  
Manufacture  
Reference

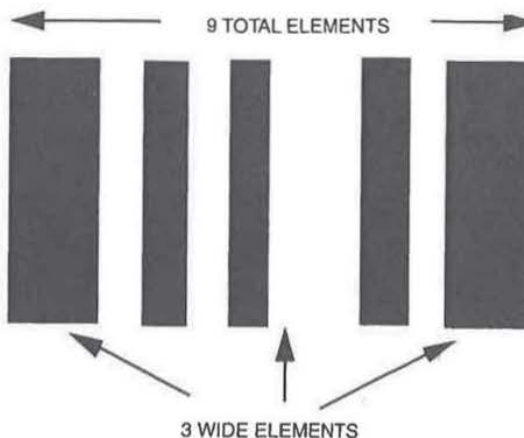
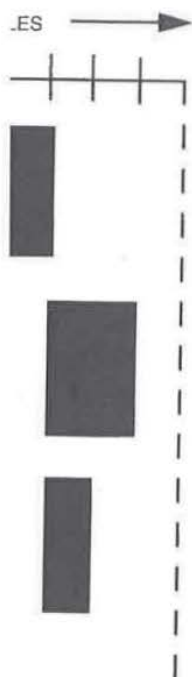


Figure 12.4 The character "A" encoded using Code 39.

Figure 12.5 shows the full encoding pattern for Code 39. Figure 12.6 shows the character string "1A" encoded using Code 39; note the use of the \* as a "start" and "stop" code in this string. Note also in Figure 12.6 the discrete nature of Code 39, with a fixed spacing between each character.

CHAR.	PATTERN	BARS	SPACES	CHAR.	PATTERN	BARS	SPACES
1	[Pattern]	1001	0100	M	[Pattern]	11000	0001
2	[Pattern]	1001	0100	N	[Pattern]	10101	0001
3	[Pattern]	11000	0100	O	[Pattern]	10100	0001
4	[Pattern]	10101	0100	P	[Pattern]	01100	0001
5	[Pattern]	10100	0100	Q	[Pattern]	00011	0001
6	[Pattern]	01100	0100	R	[Pattern]	10010	0001
7	[Pattern]	00011	0100	S	[Pattern]	11010	0001
8	[Pattern]	10010	0100	T	[Pattern]	00110	0001
9	[Pattern]	01010	0100	U	[Pattern]	10001	1000
0	[Pattern]	10110	0100	V	[Pattern]	01001	1000
A	[Pattern]	10011	0010	W	[Pattern]	11000	1000
B	[Pattern]	11001	0010	X	[Pattern]	00101	1000
C	[Pattern]	11000	0010	Y	[Pattern]	10100	1000
D	[Pattern]	10101	0010	Z	[Pattern]	01100	1000
E	[Pattern]	10100	0010		[Pattern]	10011	1000
F	[Pattern]	11100	0010		[Pattern]	10010	1000
G	[Pattern]	00011	0010	SPACE	[Pattern]	10101	1000
H	[Pattern]	10010	0010		[Pattern]	10110	1000
I	[Pattern]	01010	0010	\$	[Pattern]	00000	1110
J	[Pattern]	00110	0010	/	[Pattern]	00000	1101
K	[Pattern]	10001	0001	*	[Pattern]	00000	1011
L	[Pattern]	11001	0001	%	[Pattern]	00000	0111

\*Denotes the special start/stop code character

Figure 12.5 The Code 39 character structure. Copyright © Automatic Identification Manufacturers, Inc. 1991. This table was reproduced with permission of AIM USA from Reference 2, which is available in its entirety from AIM USA.

characters encoded using a (9,2) (n,k) encoding.

in electronic document management described in the next two sections.

Figure 12.5 is a discrete variable length code data.<sup>2</sup> Every character encoded using Code 39 (the "9" in Code 39). Of the nine elements, the "9" is wide (the "3" in Code 39) and the "0" is narrow (the "1" in Code 39). The asterisk (\*) provides for forty-four different characters (the asterisk is a special character, indicating the start and end of a string). Of the forty-three characters consisting of the following symbols: - (period) (space) \$ / + \* (asterisk) "A" encoded using Code 39. Figure



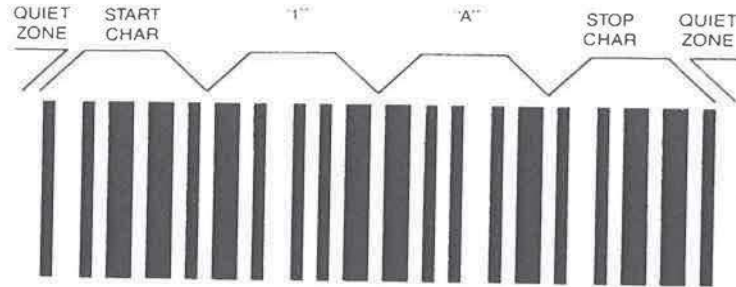


Figure 12.6 The character string 1A encoded using Code 39. Copyright © Automatic Identification Manufacturers, Inc. 1991. This figure was reproduced with permission of AIM USA from Reference 2, which is available in its entirety from AIM USA.

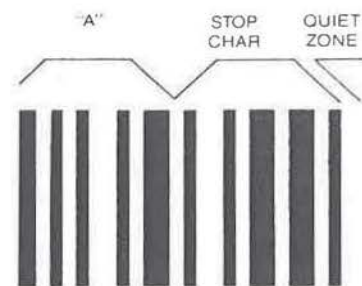
ASCII	USS-39	ASCII	USS-39	ASCII	USS-39	ASCII	USS-39
NUL	%U	SP	SPACE	0	%V		%W
SOH	\$A	!	/A	A	A	a	+A
STX	\$B	"	/B	B	B	b	+B
ETX	\$C	#	/C	C	C	c	+C
EOT	\$D	\$	/D	D	D	d	+D
ENQ	\$E	%	/E	E	E	e	+E
ACK	\$F	&	/F	F	F	f	+F
BEL	\$G	'	/G	G	G	g	+G
BS	\$H	(	/H	H	H	h	+H
HT	\$I	)	/I	I	I	i	+I
LF	\$J	*	/J	J	J	j	+J
VT	\$K	+	/K	K	K	k	+K
FF	\$L	,	/L	L	L	l	+L
CR	\$M	-	/M	M	M	m	+M
SO	\$N	.	/N	N	N	n	+N
SI	\$O	/	/O	O	O	o	+O
DLE	\$P	0	0	P	P	p	+P
DC1	\$Q	1	1	Q	Q	q	+Q
DC2	\$R	2	2	R	R	r	+R
DC3	\$S	3	3	S	S	s	+S
DC4	\$T	4	4	T	T	t	+T
NAK	\$U	5	5	U	U	u	+U
SYN	\$V	6	6	V	V	v	+V
ETB	\$W	7	7	W	W	w	+W
CAN	\$X	8	8	X	X	x	+X
EM	\$Y	9	9	Y	Y	y	+Y
SUB	\$Z	:	/Z	Z	Z	z	+Z
ESC	%A	;	%F	[	%K	{	%P
FS	%B	<	%G	\	%L		%O
GS	%C	=	%H	]	%M	~	%R
RS	%D	>	%I	^	%N		%S
US	%E	]	%J	_	%O	DEL	%T,%X,%Y,%Z

Figure 12.7 Use of Code 39 to represent the full ASCII character set. Copyright © Automatic Identification Manufacturers, Inc. 1991. This table was reproduced with permission of AIM USA from Reference 2, which is available in its entirety from AIM USA.

A bin  
each narr  
a one. Th  
representa  
narrow b  
character  
biguous  
The fo  
representa  
ASCII ch  
sentation  
Code 39 i

VALUE	
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	

Figure 12.8  
Inc. 1991. Th  
which is ava



A binary representation of Code 39 characters can be achieved by encoding each narrow element (either a bar or a space) as a zero and each wide element as a one. The letter "A" encoded as shown in Figure 12.4 would have a binary representation of 100001001 (wide bar, narrow space, narrow bar, narrow space, narrow bar, wide space, narrow bar, narrow space, wide bar). The fact that each character in this discrete coding scheme starts with a black bar leads to an unambiguous 9-bit digital representation of each character using this coding scheme.

The formal definition of Code 39 is fairly restrictive, since it is limited to representation of only 43 characters. Code 39 can be used to represent the full ASCII character set by assigning more than one character in the Code 39 representation to correspond to a single ASCII character. Figure 12.7 shows how Code 39 is used to encode the full ASCII character set.

Using Code 39. Copyright © Automatic Identification Manufacturers, Inc. 1991. This figure was reproduced with permission of AIM USA.

US-39 ASCII US-39

%V		%W
A	a	+A
B	b	+B
C	c	+C
D	d	+D
E	e	+E
F	f	+F
G	g	+G
H	h	+H
I	i	+I
J	j	+J
K	k	+K
L	l	+L
M	m	+M
N	n	+N
O	o	+O
P	p	+P
Q	q	+Q
R	r	+R
S	s	+S
T	t	+T
U	u	+U
V	v	+V
W	w	+W
X	x	+X
Y	y	+Y
Z	z	+Z
%K	:	%P
%L	:	%Q
%M	:	%R
%N	~	%S
%O	DEL	%T,%X,%Y,%Z

full ASCII character set. Copyright © Automatic Identification Manufacturers, Inc. 1991. This table was reproduced with permission of AIM USA.

VALUE	CODE A	CODE B	CODE C	BAR PATTERN	VALUE	CODE A	CODE B	CODE C	BAR PATTERN
0	SP	SP	00	2 1 2 2 2 2	60	I	I	60	3 1 4 1 1 1
1	!	!	01	2 2 2 1 2 2	61	!	!	61	2 2 1 4 1 1
2	"	"	02	2 2 2 2 2 1	62	"	"	62	4 3 1 1 1 1
3	#	#	03	1 3 1 2 2 3	63	#	#	63	2 1 1 2 2 4
4	\$	\$	04	1 2 1 3 2 2	64	NUL		64	1 1 1 4 2 2
5	%	%	05	1 3 1 2 2 2	65	SOH	a	65	1 2 1 1 2 4
6	&	&	06	1 2 2 3 1 2	66	STX	b	66	1 1 2 2 1 4
7	'	'	07	1 2 2 3 1 2	67	ETX	c	67	1 4 1 1 2 2
8	(	(	08	1 3 2 2 1 2	68	EOT	d	68	1 4 1 2 2 1
9	)	)	09	2 2 1 2 1 3	69	ENQ	e	69	1 1 2 2 1 4
10	*	*	10	2 2 1 3 1 2	70	ACK	f	70	1 1 2 4 1 2
11	+	+	11	2 3 1 2 1 2	71	BEL	g	71	1 2 2 1 1 4
12	,	,	12	1 1 2 2 3 2	72	BS	h	72	1 2 2 4 1 1
13	-	-	13	1 2 2 1 3 2	73	HT	i	73	1 4 2 1 1 2
14	.	.	14	1 2 2 2 3 1	74	LF	j	74	1 4 2 2 1 1
15	/	/	15	1 1 3 2 2 2	75	VT	k	75	2 4 1 2 1 1
16	0	0	16	1 2 3 1 2 2	76	FF	l	76	2 2 1 1 1 4
17	1	1	17	1 2 3 2 2 1	77	CR	m	77	4 1 3 1 1 1
18	2	2	18	2 2 3 2 1 1	78	SO	n	78	2 4 1 1 1 2
19	3	3	19	2 2 1 1 3 2	79	SI	o	79	1 3 4 1 1 1
20	4	4	20	2 2 1 2 3 1	80	DLE	p	80	1 1 1 2 4 2
21	5	5	21	2 1 3 2 1 2	81	DC1	q	81	1 2 1 1 4 2
22	6	6	22	2 2 3 1 1 2	82	DC2	r	82	1 2 1 2 4 1
23	7	7	23	3 1 2 1 3 1	83	DC3	s	83	1 1 4 2 1 1
24	8	8	24	3 1 1 2 2 2	84	DC4	t	84	1 2 4 1 1 2
25	9	9	25	3 2 1 1 2 2	85	NAK	u	85	1 2 4 2 1 1
26	:	:	26	3 2 2 2 1 2	86	SYN	v	86	4 1 1 2 1 2
27	;	;	27	3 1 2 2 1 2	87	ETB	w	87	4 2 1 1 1 2
28	<	<	28	3 2 2 1 1 2	88	CAN	x	88	4 2 1 2 1 1
29	=	=	29	3 2 2 2 1 1	89	EM	y	89	2 1 2 1 4 1
30	>	>	30	2 1 2 1 2 3	90	SUB	z	90	2 1 4 1 2 1
31	?	?	31	3 1 2 3 2 1	91	ESC		91	4 1 2 1 2 1
32	@	@	32	2 1 2 1 2 1	92	FS		92	1 1 1 1 4 3
33	A	A	33	1 1 1 3 2 3	93	GS		93	1 1 1 3 4 1
34	B	B	34	1 3 1 1 2 3	94	RS		94	1 3 1 1 4 1
35	C	C	35	1 2 1 3 1 3	95	US	DEL	95	1 1 4 1 1 3
36	D	D	36	1 1 2 3 1 3	96	FNC 3	FNC 3	96	1 1 4 3 1 1
37	E	E	37	1 3 2 1 1 3	97	FNC 2	FNC 2	97	4 1 1 1 1 3
38	F	F	38	1 3 2 3 1 1	98	SHIFT	SHIFT	98	4 1 1 3 1 1
39	G	G	39	2 1 1 3 1 3	99	CODE C	CODE C	99	1 1 3 1 4 1
40	H	H	40	2 3 1 1 1 3	100	CODE B	FNC 4	100	1 1 3 1 4 1
41	I	I	41	2 3 1 3 1 1	101	FNC 4	CODE A	101	3 1 1 1 4 1
42	J	J	42	1 1 2 1 3 3	102	FNC 1	FNC 1	102	4 1 1 1 3 1
43	K	K	43	1 1 2 3 3 1					
44	L	L	44	1 3 2 1 3 1					
45	M	M	45	1 1 3 1 2 3					
46	N	N	46	1 1 3 3 2 1					
47	O	O	47	1 3 3 1 2 1					
48	P	P	48	3 1 3 1 2 1					
49	Q	Q	49	2 1 3 3 1 1					
50	R	R	50	2 3 1 1 3 1					
51	S	S	51	2 1 3 1 1 3					
52	T	T	52	2 1 3 3 1 1					
53	U	U	53	2 1 3 1 3 1					
54	V	V	54	3 1 1 1 2 3					
55	W	W	55	3 1 1 3 2 1					
56	X	X	56	3 1 1 2 2 1					
57	Y	Y	57	3 1 2 1 1 3					
58	Z	Z	58	3 1 2 3 1 1					
59	[	[	59	3 3 2 1 1 1					

Figure 12.8 Code 128 symbology. Copyright © Automatic Identification Manufacturers, Inc. 1991. This table was reproduced with permission of AIM USA from Reference 3, which is available in its entirety from AIM USA.



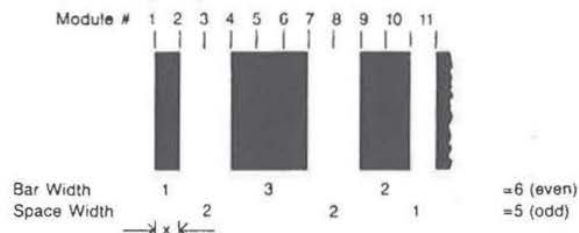
## Code 128

Code 128 is a variable length, continuous code that is an (11,3) (n,k) code.<sup>3</sup> The character structure is shown in Figure 12.8. There are three different character sets encoded using the Code 128 symbology. Encoded character strings are preceded by a "start" character that indicates which of the three character sets is used for the subsequent characters. The three character sets are referred to as "Code A", "Code B", and "Code C." Start characters can be inserted into a variable length character string at any point, so that shifts between the three character sets can be implemented within a single character string.

A SHIFT character is also available. Use of the shift character changes the code representation from A to B or from B to A for the next individual character following the SHIFT character.

Check characters are used to insure correct interpretation of scanned bar code information in Code 128. The check character is inserted prior to the STOP character in an encoded character string. The check character has a numerical value determined by the element pattern in the encoded string preceding the check character. If the value of the check character differs from a numerical calculation based on the interpreted character string, the scan is not valid.

### Character 1 (one)



### STOP Character

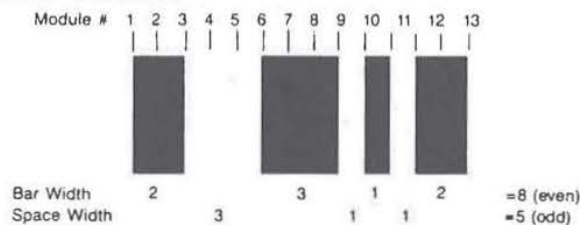


Figure 12.9 Basic character structure for Code 128. (Top) representation of the character 1, (Bottom) the STOP character. Copyright © Automatic Identification Manufacturers, Inc. 1991. This figure was reproduced with permission of AIM USA from Reference 3, which is available in its entirety from AIM USA.

Figure  
Charact  
before  
from R

Fig  
There  
three s  
Fig  
charact  
bar cod  
charact

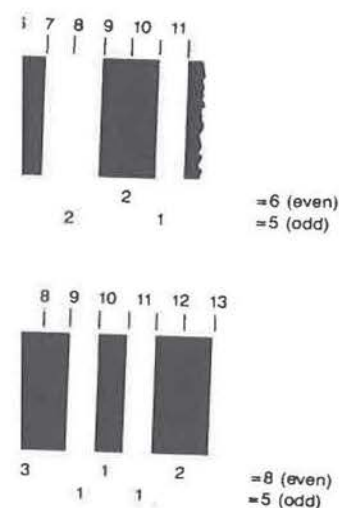
## USE

The us  
ment n  
bar cod  
incorpor  
The ba  
scanne  
taining  
method  
interpre  
has occ  
applica

Uni  
code li  
establi  
unique  
code id

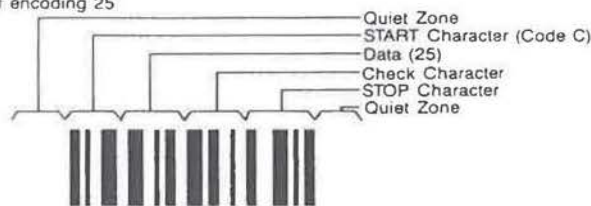
ous code that is an  $(11,3)$   $(n,k)$  code.<sup>3</sup> The  
: 12.8. There are three different character  
ology. Encoded character strings are pre-  
ates which of the three character sets is  
he three character sets are referred to as  
tart characters can be inserted into a vari-  
it, so that shifts between the three charac-  
ngle character string.  
2. Use of the shift character changes the  
n B to A for the next individual character

correct interpretation of scanned bar code  
character is inserted prior to the STOP  
g. The check character has a numerical  
ern in the encoded string preceding the  
k character differs from a numerical cal-  
ter string, the scan is not valid.



de 128. (Top) representation of the  
yright © Automatic Identification  
roduced with permission of AIM USA  
irety from AIM USA.

USS-128 Symbol encoding 25



**Figure 12.10** Encoded character string for the number 25 using Code 128. The START Character indicates that the Code C subset applies. Note presence of check character before the STOP character. This figure was reproduced with permission of AIM USA from Reference 3, which is available in its entirety from AIM USA.

Figure 12.9 shows the basic structure of encoded characters using Code 128. There are eleven modules, and each character is comprised of three bars and three spaces. The STOP character is also shown in Figure 12.9.

Figure 12.10 shows the full Code 128 encoding of the number 25. The start character denotes that the Code C is being used for the next character string. The bar code representation of the numerical value 25 is next, followed by a check character and the STOP character.

## USE OF BAR CODE TECHNOLOGY IN ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS

The use of bar code technology is becoming more common in electronic document management systems. Electronic document management systems may read bar code information in several different ways. One common method involves incorporation of bar code scanners as peripheral units on paper scanner units. The bar code device scans the paper as it is fed through the scanner, and the scanner server receives both the scanned document image and a data record containing the alphanumeric information read by the bar code scanner. Another method involves use of computer software on the scanner server for bar code interpretation, operating from the scanned document page image after scanning has occurred. Examples of the use of this technology in document processing applications include the following:

**Unique Identifier for Indexing.** In one common application, preprinted bar code labels are attached to pages prior to scanning. The bar code labels can establish unique index identifiers for individual pages, or can be used to uniquely index a series or batch of pages at a higher level—for example, bar code identifiers can be used to denote the first page of a new document.



**Form Type.** Imprinted bar codes can be used to uniquely identify a particular type of form in electronic document management systems. Applications software can be used to search for bar code signatures at specific locations on scanned page images. The decoded bar code information can then be used by applications software for higher level processing. For example, some applications may involve scanning many thousands of forms, where the number of form types may be a relatively small fixed number (there may only be a few types of forms involved in a specific application). If that is the case, the pre-printed information on the form itself will be redundant after scanning—all scanned form images will contain the same information which has been preprinted on the form. The actual information content may involve only a small percentage of the scanned material, corresponding to the information that has been entered on the form by an individual. If the form type can be identified automatically, it is possible to extract the segments of the image containing the information content and to discard the remainder of the scanned image. This automated forms identification process can result in a significant reduction in data volume within an electronic document management system since only the small image segments containing information are retained for further use.

**Batch Identification.** In some applications, document material may be divided up into batches for scanning. Each batch may contain document material that forms a logical unit within the application, or a batch might simply denote the date, time and operator involved in scanning the material. In either case, use of a bar code label on the first scanned page, or use of bar code on a separate control page that is scanned prior to the first document page, can be used to provide a unique index for the batch of material. Control pages may be inserted periodically into batches of scanned material for a variety of reasons in some applications. The applications software identifies the unique bar code, and structures the scanned data logically.

## References

1. Palmer, R.C., *The Bar Code Book*, Helmers Publishing, Inc., Peterborough, NH, Second Edition, 1991.
2. Automatic Identification Manufacturers, *USS-39 Uniform Symbology Specification*, Pittsburgh, PA, 1986.
3. Automatic Identification Manufacturers, *USS-128 Uniform Symbology Specification*, Pittsburgh, PA, 1986.

Sol

RE

Black, D.I

Kaltohoff,  
OH, StrateD'Alleyra  
PublicatioHall, G.A  
McGraw ID'Alleyra  
New YorkBeale, S.,  
Press, 198

REFI

Castlemar  
Hall, 1975