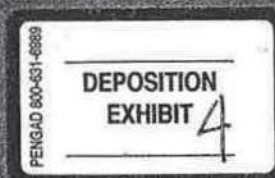


# COMPUTERIZED DOCUMENT IMAGING SYSTEMS

TECHNOLOGY AND APPLICATIONS

NATHAN J. MULLER



## **Computerized Document Imaging Systems: Technology and Applications**

For a complete listing of the *Artech House Telecommunications Library*,  
turn to the back of this book.

# **Computerized Document Imaging Systems: Technology and Applications**

Nathan J. Muller  
Illustrations by Linda Lee Tyke

Artech House  
Boston • London



Library of Congress Cataloging-in-Publication Data

Muller, Nathan J.

Computerized document imaging systems: technology and applications / Nathan Muller  
Includes bibliographical references and index.

ISBN 0-89006-661-2

1. Business records—Management—Data processing. 2. Imaging systems. 3. Documents in optical storage. 4. Information storage and retrieval systems—Business. I. Title.

HF5736.M84 1993  
651.5'0285—dc20

92-37800  
CIP

British Library Cataloguing In Publication Data

Muller, Nathan J.

Computerized Document Imaging Systems: Technology and Applications

I. Title

006.4

ISBN 0-89006-661-2

© 1993 ARTECH HOUSE, INC.  
685 Canton Street  
Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

International Standard Book Number: 0-89006-661-2  
Library of Congress Catalog Card Number: HF5736.M84 1993

10 9 8 7 6 5 4 3 2 1

*To my son, Matthew*

## *Contents*

Introduction	xv
Part I Basic Principles	
Chapter 1 Document Imaging Concepts	3
1.1 Introduction	3
1.2 Office Document Imaging Systems	5
1.2.1 General Business Systems	5
1.2.2 Image Processing Systems	5
1.2.3 Engineering Drawing Systems	6
1.2.4 Mapping Systems	6
1.3 Imaging Functions	6
1.3.1 Scanning and Indexing	7
1.3.2 Storage and Retrieval	7
1.3.3 Compression	8
1.3.4 Workflow Automation	8
1.3.5 Office Automation	8
1.4 Imaging System Components	8
1.5 Market Players	9
1.6 Complementary Technologies	10
1.6.1 Electronic Data Interchange	10
1.6.2 Multimedia	11
1.6.3 Videoconferencing	11
1.6.4 Facsimile	12
1.7 Network Requirements	12
1.8 Selection Criteria	14
1.9 Impact of Open Systems	14
1.10 Determining Needs	15
1.11 Conclusion	16

Chapter 2	Information Management	19
2.1	Introduction	19
2.2	Text Search and Retrieval	19
2.2.1	Key-Word Indexing	20
2.2.2	Full Text Retrieval	21
2.2.3	Boolean Search	21
2.2.4	Automatic Clustering by Topic	22
2.2.5	Concept-Based Retrieval	22
2.3	Structured Query Language	23
2.4	Hypertext	26
2.5	Compound Documents	28
2.6	Graphics Files	31
2.7	The Future of Image Retrieval	32
2.8	Conclusion	33
Chapter 3	Imaging Applications	35
3.1	Introduction	35
3.2	Application Overviews	35
3.3	Technical Documentation Application	38
3.3.1	The Problems	39
3.3.2	The Solutions	40
3.4	Credit Card Application	42
3.4.1	The Problems	42
3.4.2	The Solutions	45
3.5	Accounts Payable Application	48
3.5.1	The Problems	48
3.5.2	The Solutions	48
3.6	Insurance Processing Application	52
3.6.1	The Problems	52
3.6.2	The Solutions	53
3.7	Loan Services Application	54
3.7.1	The Problems	54
3.7.2	The Solutions	55
3.8	Conclusion	55
Part II	Imaging System Platforms	
Chapter 4	Desktop Imaging Systems	59
4.1	Introduction	59
4.2	Functional Components	60
4.2.1	Image Capture	60
4.2.2	Indexing	65



---

4.2.3	Compression	65
4.2.4	Storage	68
4.2.5	High-Resolution Displays	71
4.2.6	Management	73
4.3	Conclusion	73
Chapter 5	Considerations for LAN-Based Imaging	75
5.1	Introduction	75
5.2	LAN Approaches	75
5.3	LAN Characteristics	77
5.3.1	Topologies	78
5.3.2	Access Methods	78
5.3.3	Media Types	79
5.4	Types of LANs	80
5.4.1	Ethernet	80
5.4.2	Token Ring	82
5.4.3	ARCnet	83
5.4.4	StarLAN	85
5.4.5	FDDI	85
5.4.6	SDDI	86
5.5	The Role of Hubs	87
5.6	The Role of Servers	88
5.7	Conclusion	89
Chapter 6	Considerations for Mainframe-Based Imaging	91
6.1	Introduction	91
6.2	Role of the Mainframe	94
6.3	Mainframe Image System Components	94
6.3.1	Workstations	94
6.3.2	Scanning Equipment	95
6.3.3	Printers	97
6.3.4	Communications Controllers	98
6.3.5	Storage Facilities	98
6.4	Implementation Issues	103
6.5	Conclusion	104
Part III	Vendor Implementations	
Chapter 7	IBM Imaging Systems	107
7.1	Introduction	107
7.2	Imaging System Offerings	107
7.3	ImagePlus Overview	108
7.3.1	Folder Application Facility	108

	7.3.2	Object Distribution Manager	111
	7.3.3	Object Access Method	111
	7.4	Document Entry	113
	7.5	Workflow	114
	7.6	Folder Management	116
	7.7	Convenience Features	118
	7.8	Supervisor Capabilities	119
	7.9	Online Help	119
	7.10	ImagePlus/2 Application Facility/2	120
	7.10.1	System Administration Program	120
	7.10.2	Image Processing Program	120
	7.11	Consulting Services	121
	7.12	Education Services	121
	7.13	Conclusion	122
Chapter 8		Unisys InfoImage System	123
	8.1	Introduction	123
	8.2	Product Overview	124
	8.3	Software Components	126
	8.3.1	Folder Application Software	126
	8.3.2	Director	129
	8.3.3	Database Software	131
	8.3.4	Event Manager	132
	8.4	Hardware Components	133
	8.4.1	Server	134
	8.4.2	Optical Disk Drives	135
	8.4.3	Workstations	135
	8.4.4	Communications Boards	137
	8.4.5	Scanners	137
	8.4.6	Printers	138
	8.5	Conclusion	138
Chapter 9		Hewlett-Packard Advanced Image Management System	139
	9.1	Introduction	139
	9.2	System Overview	139
	9.3	AIMS DataManager	141
	9.3.1	Free Text Search	143
	9.3.2	Optical Storage Clustering	143
	9.3.3	DataManager Modules	144
	9.4	AIMS MediaManager	146
	9.5	AIMS JukeboxManager	148
	9.6	Print Server	149
	9.7	Fax Server	150

9.8	Applications Integration	150
9.8.1	HP 3000 and HP 9000 Integration	151
9.8.2	IBM 3270 Integration	151
9.9	Conclusion	151
Chapter 10	DECimage EXpress	153
10.1	Introduction	153
10.2	Development Background	154
10.3	System Overview	155
10.3.1	DECwindows Motif	157
10.3.2	Compound Document Architecture	158
10.3.3	Document Interchange Formats	159
10.4	CDA Applications	161
10.5	Network Applications Support	162
10.6	DECimage Application Services	163
10.7	Application Development Environment	164
10.8	Other DECimage Offerings	165
10.8.1	All-in-1 Image Software	165
10.8.2	DECimage Character Recognition Services	165
10.8.3	SMARTFOLDER Imaging	165
10.8.4	DECimage Scan Software	166
10.8.5	DECcamera	166
10.8.6	Fax Network Gateway Server	166
10.8.7	Storage	166
10.9	Third-Party Relationships	166
10.9.1	ImageNow Software	166
10.9.2	LaserView Express Software	167
10.9.3	Epic/Workflow Software	168
10.9.4	PixTex Product Family	168
10.10	Conclusion	168
Chapter 11	FileNet Document Imaging System	171
11.1	Introduction	171
11.2	System Overviews	171
11.2.1	WorkFlo Business System	171
11.2.2	FolderView	173
11.3	Image System Components	174
11.4	WorkShop Development Environment	176
11.4.1	Forms Development	176
11.4.2	Forms Build	176
11.4.3	Script Development	183
11.4.4	Dynamic Data Exchange	184
11.5	Support Services	187

11.6	Conclusion	188
Part IV Wide-Area Networking		
Chapter 12	Considerations for Imaging Over WANs	191
12.1	Introduction	191
12.2	Transmission Facilities	192
12.2.1	Dial-Up and Analog Leased Lines	192
12.2.2	Packet-Switched Services	192
12.2.3	Digital Data Services	193
12.2.4	Fractional T1	193
12.2.5	Integrated Services Digital Network	194
12.2.6	Dedicated T1 Lines	194
12.2.7	Frame Relay	194
12.2.8	T3	195
12.2.9	Fractional T3	196
12.2.10	Switched Multimegabit Data Services	196
12.2.11	Broadband ISDN	197
12.3	LAN Interconnection Devices	198
12.3.1	Bridging Methods	198
12.3.2	Routers	201
12.4	Bandwidth on Demand	202
12.4.1	Inverse Multiplexing	203
12.4.2	LAN Interconnection	204
12.5	Conclusion	205
Part V Related Integration Areas		
Chapter 13	Videoconferencing and Image Transfer	209
13.1	Introduction	209
13.2	Applications	211
13.3	Image Transfer Capabilities	212
13.4	Wide-Area Video Transmission	215
13.4.1	Very Small Aperture Terminals	215
13.4.2	Microwave	217
13.4.3	Fiber Optics	217
13.4.4	Integrated Services Digital Network	219
13.4.5	Local-Access Channels	220
3.5	Conclusion	223
Chapter 14	Multimedia Imaging	225
14.1	Introduction	225
14.2	Workstation Requirements	226



14.3	Compression Methods	229
14.3.1	Joint Photographic Experts Group	229
14.3.2	Motion Picture Experts Group	230
14.3.3	Px64	230
14.3.4	Digital Video Interactive	230
14.3.5	Fractal Compression	232
14.4	Real-World Business Applications	232
14.4.1	Multimedia Kiosks	232
14.4.2	Other Uses for Kiosks	233
14.4.3	Kiosks Yesterday and Today	234
14.4.4	Applications Development	234
14.4.5	Network Integration	235
14.5	Multimedia with E-Mail	237
14.6	Multimedia and Hypertext	237
14.7	Multimedia Authoring Tools	237
14.8	Multimedia Standards	238
14.9	Conclusion	240

## Part VI Planning and Services

Chapter 15	Planning and Implementation Issues	243
15.1	Introduction	243
15.2	Organizational Issues	244
15.3	Planning Considerations	245
15.3.1	Indexing Versus Workflow	245
15.3.2	Understanding Processing Requirements	245
15.3.3	Levels of Implementation	247
15.3.4	Role of Operating System	249
15.3.5	Topology Considerations	249
15.3.6	Impact on Existing Networks	250
15.3.7	Applications Integration	253
15.4	Support Issues	254
15.4.1	System Reliability	254
15.4.2	Systems Integration	255
15.4.3	Training	256
15.5	Conclusion	256
Chapter 16	Professional Services for Imaging Implementation	259
16.1	Introduction	259
16.2	Productivity Issues	260
16.3	Professional Services Overview	262
16.3.1	Predecision Services	264
16.3.2	Preinstallation Services	265

16.3.3	Installation and Implementation Services	266
16.3.4	Postinstallation Services	267
16.3.5	Life Cycle Services	267
16.3.6	Value-Added Systems Integration	267
16.4	Teaming Arrangements	268
16.5	Other Vendor Services	269
16.5.1	Wang Laboratories	269
16.5.2	IBM	270
16.5.3	DEC	271
16.6	Vendor-Provided Service: Pro and Con	271
16.7	Conclusion	272
Appendix	Specialized Imaging Systems	275
	About the Author	279
	Glossary	281
	Bibliography	291
	Index	293

## *Introduction*

Corporate America's huge investment in personal computers has done little to ameliorate white-collar productivity, except to generate more paper. Yet there is still hope for improvement. Because of advances in technology and the consequent lower cost inherent in mass production, more companies are willing to consider computerized document imaging systems for improving business processes and workflow dynamics.

Companies of many types and sizes are reaping efficiency improvements and productivity increases through implementations of document imaging systems that simplify and reduce access time and enhance document integrity. Users can retrieve documents much faster from electronic systems than is possible with traditional, paper-based methods. Also, access to documents can be effected by a variety of methods involving content-based retrieval techniques. Document imaging systems can minimize paper use and ultimately reduce the amount of space required for storage.

Contributing to the growing popularity of document imaging systems are decreased hardware costs and the standardization of interfaces between major components. These advances have resulted in a move away from proprietary hardware and toward general-purpose hardware. Standards enable users to run image applications on nondedicated systems and across different hosts and networks. *Application programming interfaces* (APIs) and other development tools permit existing programs to be image-enabled.

Some vendors have introduced imaging systems that allow users to build applications that run on mainframes, microcomputers, and *local-area networks* (LANs). Other vendors offer script- and object-oriented languages that facilitate the building of workflow applications that work under *Microsoft disk operating system* (MS-DOS) and *Microsoft Windows* (MS-Windows).

### **Benefits of Imaging**

Modern imaging technology allows a document to be captured digitally via scanning, converted into a bit-mapped image, indexed, compressed, and stored in a database. The image can then be retrieved and distributed for processing.

Workflow software, meanwhile, has advanced to deliver a multitude of benefits that can result in lower operating costs, less exposure to litigation, and higher revenue potential for many businesses.

The first benefit is better document access. Because documents are stored electronically, and retrieval is accomplished via key field values, documents can be retrieved very quickly. Documents stored on line on optical or magnetic disk can be retrieved virtually instantaneously, and those on jukeboxes (which are sometimes described as "near on line") can be retrieved in under a minute. This contrasts with waiting up to hours to retrieve information from a file cabinet, or up to days to retrieve information that has been archived in a warehouse.

Furthermore, with content-based retrieval a file can be accessed in different ways, such as by case number, date, or type. Documents can be stored as both text and image and retrieved based on any word or combination of words. And, because the file is like any other computer record, it can be accessed concurrently by multiple workstation operators.

A second major benefit of computerized document imaging systems is vastly improved document integrity. Because files are stored electronically, they are less likely to get lost or physically damaged. Relational database features reconcile document changes and carry them forward to other databases. Additionally, it is easy to create backups and store them off-site for added security.

Reduction in the amount of paper used is a third major benefit. Fewer paper copies are needed because the document can be stored electronically. The physical space for paper is also greatly reduced—a single 5.25-in optical disk (approximately 600 MB) holds a quantity of images equivalent to all the papers stored in a four-drawer file cabinet.

It is not uncommon for companies to cost-justify imaging technology based on a short payback period. Tangible savings from eliminating microfilm, reducing the head count of clerical staff, and saving storage space can result in payback periods of 18 months or less. When intangibles such as better response, improved security, and fewer lost documents are added in, the payback period can be under one year.

### **Workflow Automation and Management**

The biggest payoff comes from using images to automate and manage workflow. Workflow processing actively routes documents through a system, using rules that reflect the decision criteria that is used to process the documents. Instead of just dumping all information into an "in-basket," logical queues of documents are established that enable workstation operators to obtain the next available document for processing. This results in streamlined processing of documents, speedier distribution that requires fewer people, and more productivity from the people who must process the information.

For example, incoming documents are scanned into the system, and then they are indexed by the claim number, name, form type, and scan date. All documents for the same case can be grouped in an electronic file folder. When a file folder is opened, the whole folder or one or more individual documents are routed immediately to the appropriate



queue. Other information might reside in a mainframe database and can be integrated into the document via terminal emulation. In fact, multiple mainframe sessions can be established, each displayed in a separate window on the image workstation.

When a document processing step is completed, the file is placed in the appropriate logical queue, to be forwarded to the next processing step. No physical transfer need take place. If further work is needed, the document can be sent back to the person who previously handled it, or can be held in suspense until the additional information is obtained.

With fewer people required to process documents, compared to a paper-based system, an organization can be more responsive to the needs of its customers or clients. In the past, important documents might have taken days to get from the mailroom to the right person's desk. That wait has been replaced by the capability to transfer document images instantaneously.

Another benefit of imaging systems is the ability to manage the flow of work through the office. Supervisors can easily monitor the status of documents and files to ensure that high-priority cases are handled expeditiously. The results are more productive workers and happier clients.

Some industry observers view this new way of document handling as signaling a paradigm shift from data management to work management. According to this view, imaging systems are changing not only the kind of information that moves within organizations, but how it moves as well.

Workflow systems are particularly successful when designed to allow the process to be altered to match the technology. This radical change in information systems philosophy is called *business process redesign*, or *business re-engineering*. Instead of simply computerizing an existing manual system, a business must determine the goal of the manual process. Then the business must look for a way to achieve that goal in an efficient, cost-effective manner, even if it requires altering the current business practice.

Projects in which business processes have been altered to take advantage of new technologies often result in the biggest productivity increases. However, caution should be taken to prevent sudden and sweeping changes. After all, when people perceive a threat to job security, they might rebel against change and jeopardize success, either through inaction or deliberate means.

In fact, the organizational problems are often more difficult to solve than the technical ones. For example, a large insurance company wanted to implement an imaging system in its field offices, but was concerned that the staff would feel threatened and resist contributing to the design. The company solved this problem by selecting one field office for a pilot and guaranteeing job security regardless of the outcome of the new system. This not only alleviated the immediate problem of staff concerns, the staff's increased participation also ensured the long-term success of the project.

### Integration Issues

Any image system implementation project requires that images be integrated with existing applications and with record-oriented data that might already exist in a corporate database.

Accordingly, an important factor in the purchasing decision is the ability of the product to integrate with existing equipment and software. The strong requirement for integration is also one of the major forces pushing vendors toward "open systems" built on industry-standard platforms.

At the applications level, integration is achieved by setting up image windows and database windows on the same screen. Data can then be transferred from the database window to the image window. This type of data transfer can be done in a variety of ways, including through the use of MS-Windows' linkage mechanism, *dynamic data exchange* (DDE).

A much better approach than using DDE is to use cooperative processing to build applications that span microcomputers, minicomputers, and mainframes. Portions of an application can run on any machine in a network, and data from a variety of databases can be linked with images. To provide easier integration of images with existing applications, APIs are used, allowing images to be accessed from standard programming languages.

Another reason for integration is that users do not want to bring in a new *database management system* (DBMS) just to support the imaging system. Not only does this increase the cost of the imaging system, it also makes support of the system more difficult. The solution that many vendors have chosen is to allow their image filing and retrieval systems to work with a variety of *relational database management systems* (RDBMSs), including Microsoft/Sybase SQL Server and Oracle, from Oracle Corp.

One possible result of the need for integration and the push to standard systems might be the absorption of image management into the standard data processing mix. While the interfaces are now mostly nonstandard, that is changing. As systems builders learn the new design techniques necessary to succeed with imaging applications, those techniques will become part of any designer's skill repertoire.

### Technical Issues

Implementing an imaging system is a difficult technical task. The problem with images stems from the size of the bit map. For example, a scanned 8.5-in by 11-in text document can produce a file size of over 1 MB. Even with compression, the file is still between 50 KB and 100 KB. For color documents, these storage requirements may be 10 or 20 times larger.

This large size affects every aspect of image manipulation. Primarily, the large size makes file compression a requirement. Compression occurs after the image is scanned, and decompression occurs when it is displayed or printed. Until recently, compression and decompression were done in hardware, because of the amount of computation required. Add-on boards were required for adequate performance. But these boards added significantly to the cost of each display station. The trend today is toward software-based compression, which on a fast 80386 or 80486 *personal computer* (PC) provides a comparable level of performance—and at far lower cost.

Integration with DBMSs is also a problem. Even when the DBMS supports a *binary large object* (BLOB) data type, in which image data can be stored, most DBMSs do not do



all that is necessary to support images. For example, the DBMS must permit the image to be physically stored separately from the rest of the record, so that it can be put on optical media, which provides more long-term storage capacity. Furthermore, when a record is retrieved, the image buffers need to be separate from the record buffers, otherwise the size of the image will quickly fill the allocated space. Some DBMS vendors are adding these and other capabilities to facilitate the treating of images as just another data type.

Another potential problem area is networking. A large imaging application might require shipping many compressed images of 100 KB or more across the network. While many LANs can be configured to accommodate imaging applications, sending images over the slower *wide-area network* (WAN) results in unacceptable performance, in which case a wide-bandwidth solution is required. Such solutions include private T1 lines operating at 1.536 Mbps or emerging public network offerings, such as frame relay or *switched multimegabit data services* (SMDS).

This book covers these and other important issues related to the effective design and implementation of computerized document imaging systems. The book progresses from basic concepts to applications, to vendor implementations, to planning. As such, it is of value to those who need a central source of introductory information, as well as to those who are well along in planning or implementing document imaging systems and can use further input into their decision-making processes. The book is written with a minimum of technical jargon, yet the subject matter is covered in sufficient detail to appeal to those who are more technically oriented.

Although microfilm and microfiche are still important imaging technologies, the projected market growth for these systems is only 5% per year. There are also specialized imaging systems for medical applications, such as *magnetic resonance imaging* (MRI) and *computerized tomography* (CT). There are highly sophisticated imaging systems used in oil and mineral exploration and environmental applications. Likewise, there are many experimental imaging technologies, such as virtual reality and three-dimensional holography. These specialized topics are beyond the scope of this book, especially since there are many fine books already available that very capably treat these subjects. Accordingly, the focus of this book is the computerized document imaging systems used in today's office environment.

This book is organized into six parts: basic principles, imaging system platforms, vendor implementations, wide-area networking, related integration areas, and planning and services.

Part I describes the basic principles—introducing the reader to concepts, information management issues, and applications—and lays the foundation for more detailed discussions of document imaging.

Chapter 1 provides an overview of document imaging concepts, introducing the reader to the basic categories of office document imaging systems and the various discrete processes that comprise the imaging system function. These processes include scanning and indexing, storage and retrieval, compression, and workflow and office automation. Also covered in this chapter are overviews of imaging system components, complementary technologies, and networking requirements. The chapter ends with a discussion of the

impact of open systems on imaging and some of the considerations that go into needs assessment.

Chapter 2 provides a general discussion of information management. With this background, the reader will have a better understanding of how image documents are organized for storage, search, and retrieval. Among the key concepts covered in this chapter are Boolean searches, *structured query language* (SQL), hypertext, and various document architectures.

Chapter 3 provides in-depth descriptions of selected office applications of document imaging, including processing technical documentation, credit card transactions, accounts payable transactions, insurance claims, and loan origination documents. Other applications are briefly mentioned as well. The intent of this chapter is to highlight some practical applications of document imaging and to stimulate readers to imagine how document imaging might be applied to their organizations to streamline workflows, alleviate paper logjams, and increase office productivity.

Part II describes the major platforms on which document imaging systems are built: microcomputers (desktops), LANs (departmental systems), and mainframes (enterprise-wide systems).

Chapter 4 gives an overview of a category of document imaging products that are commonly referred to as *desktop imaging systems*. This type of imaging system is aimed at small businesses or work groups within the departments of large organizations. The chapter describes a range of affordable technologies and system components that come together in the low-volume office environment.

Chapter 5 describes LAN-based imaging systems and some of the considerations that make for successful implementation. This chapter explains some of the popular LAN approaches, along with the associated advantages and disadvantages. Since the choice of a LAN can have a bearing on imaging system performance, this chapter includes a discussion of the various LAN topologies. A synopsis of the popular LAN types is also provided, including Ethernet, token ring, StarLAN, ARCnet and FDDI. In addition, the roles of hubs and servers on the LAN are described.

Chapter 6 describes mainframe-based imaging. Despite the much publicized migration of mainframe applications to the distributed computing environment characterized by LANs, many companies still have a large investment in mainframe computers and are seeking to leverage that investment by adding applications such as imaging. This chapter explores the role of the mainframe within the context of document imaging applications and highlights the associated system components. For the most part, the types of system components used in the mainframe environment are the same as those used in desktop and LAN-based imaging systems; they differ only in their capacity to support a high volume of transactions. Of course, there are equipment types, such as *front-end processors* (FEPs), communications controllers, and certain types of storage facilities, that are unique to the mainframe environment. These too are covered.

Part III is a series of chapters that focus on the document imaging products of specific vendors. The selection of vendors was dictated by the willingness of the various companies to provide technical documentation and other proprietary information that



adequately described the operation of their systems. Accordingly, the cooperation of IBM Corp., Unisys Corp., Hewlett-Packard Co., *Digital Equipment Corp.* (DEC), and FileNet Corp. was instrumental in the development of the core content of this book. If only one of these vendors declined to participate, I would have considered this book seriously deficient. So I am very appreciative of their assistance.

Chapter 7 describes IBM's ImagePlus system. Although IBM offers a version of ImagePlus for all three processing platforms—mainframe, midrange, and microcomputer—this chapter focuses on the SAA ImagePlus MVS/ESA system, which supports high-volume operational and production applications.

Chapter 8 discusses Unisys' InfoImage system, highlighting the major hardware and software components and describing how the components work together to provide customers with a comprehensive imaging solution.

Chapter 9 describes Hewlett-Packard's *Advanced Image Management System* (AIMS). In addition to providing information on basic system operation, this chapter highlights the software modules that are used to support specific imaging system functions, as well as the various tools that are available for applications integration.

Chapter 10 describes DECimage EXpress, including its development history, functionality, and applications-development environment. Options that can be added to DECimage EXpress are discussed, as well as third-party offerings that are based on DECimage EXpress and can be brought into DECimage EXpress to expand its functionality or provide links into DECimage EXpress.

Chapter 11 describes the document image products of FileNet Corp., specifically, its WorkFlo Business System and FolderView offerings. This chapter also provides information on how applications are developed using object-oriented and script-oriented programming methods.

Part IV is concerned with wide-area networking. Its single chapter, Chapter 12, provides a survey of transmission facilities and services that are available for users of public or private networks. Each type of facility and service is reviewed according to its ability to support the transmission of document images. The facilities described include analog leased lines, T1 and Fractional T1, and T3 and Fractional T3. The services described include dial-up, packet-switched, *digital data services* (DDS), the *integrated services digital network* (ISDN), frame relay, and *switched multimegabit data services* (SMDS). Also covered in this chapter is *inverse multiplexing*, which is an economical way of supporting image applications by locally assembling bandwidth increments of 56/64 Kbps into higher-speed digital pipes, rather than paying high monthly local-access fees for dedicated T1 lines.

Part V contains two chapters on topics that, increasingly, will have a relationship with document imaging: videoconferencing and the emerging set of technologies that comprise multimedia.

Chapter 13 provides an introduction to videoconferencing and, along the way, describes how document imaging can be integrated into conferencing sessions.

Chapter 14 describes multimedia technologies and applications, as well as how document imaging can be integrated into multimedia presentations.

Part VI addresses the formative stages of document imaging system implementation.

Chapter 15 deals with a variety of issues that must be addressed to ensure the successful planning, acceptance, and implementation of a document imaging system. These issues include personnel, political, and technical concerns, as well as support requirements.

Chapter 16 focuses on the professional services that vendors offer to help customers assess the need for document imaging systems. The chapter evaluates the potential performance of these companies and the potential return on investment for the services. The chapter focuses on the professional services offered by Unisys Corp., although most of the large imaging system vendors also offer similar types of services. Unisys was selected as the illustration because it was the most willing of any other vendor to discuss this type of service in great detail.

The information contained in this book, especially as it relates to specific vendors and products, is believed to be accurate at the time it was written and is, of course, subject to change with continued advancements in technology and shifts in market forces. The mention of specific products and vendors is for illustration purposes only and does not constitute or imply an endorsement of any kind by the author or the publisher.

Nathan J. Muller

***PART I***  
***BASIC PRINCIPLES***

## *Chapter 1*

### *Document Imaging Concepts*

#### 1.1 INTRODUCTION

In the United States alone, 95% of the more than three trillion documents currently stored in corporate offices still exist in paper form, with the number growing by one billion pages daily. The cost of storing these documents is quite substantial. A typical office in a Manhattan building, for example, contains ten file cabinets, each capable of holding 10,000 pieces of paper, for a total of 100,000 pieces of paper. At \$70 a square foot, this translates into \$14,000 a year in rent, or \$0.14 a page. Assuming that each piece of paper will be copied at least once during the year adds another \$0.03 per page, or \$3,000 a year. A filing clerk to maintain the files and make the copies adds \$25,000 per year in salary and benefits. The total cost associated with storing information in only ten filing cabinets reaches \$42,000 a year or \$2.38 per page.

There are other expenses associated with keeping paper documents that drive the cost of doing business even higher. Not only is a lot of paper wasted, it is also lost. Various estimates put the average cost of tracking down a lost or misfiled document at \$120. The average executive spends one month per year waiting for the documents he or she needs. And then there is the cost of mailing documents, many of them by expensive overnight services. Implementing security procedures for sensitive documents also adds to the cost of maintaining paper documents.

Dealing with paper documents has been a lingering problem for most businesses. The first document imaging systems were based on microfilm, which came into widespread use in the late 1960s. Microfilm offered numerous benefits over paper filing systems, including easier and faster access to documents, improved file integrity, better security, longer document life, and cost savings based on space savings and productivity gains. But, while microfilm offered incremental improvements, it fell short of providing the means to efficiently process and retrieve documents, especially as advances in computer technology during the 1970s and 1980s greatly increased the output of paper. In addition, microfilm did not solve the persistent problem of documents lost because of data entry errors. Document losses of 30% were not uncommon.



Moreover, microfilm required that documents be sent off-site to be filmed, a process that typically took two weeks, during which time companies were left without access to their files. Retrieval was a mechanical task that was often as time-consuming as rummaging through documents stored in file cabinets.

According to a 1991 report by the *Association for Information and Image Management* (AIIM), although microfilm still accounts for about half the image management market, it is under intense competitive pressure from optical disk technology. Consequently, the market growth for microfilm has all but stopped, at a time when the market for optical disks is taking off. AIIM predicts that microfilm will grow at an annual rate of only 5% from now until 1995, compared with a 47% growth rate for electronic image management using optical disks.

Microfilm system vendors like Eastman Kodak, Bell & Howell, and 3M sensed a limited future for their products, correctly assessing that most companies would bypass microfilm and store their records on a more compact, readily accessible online medium, such as optical disk. From this defensive posture, these vendors began investing substantial research and development resources in digital imaging systems—systems that would largely mimic their existing products but replace microfilm and microfiche with optical disk technology.

Of note is the new strategy of some vendors of providing customers with the means to facilitate the migration from microfilm to electronic document storage. Kodak's Imagelink Scanner/Microimager 990S, for example, can capture as many as 120 documents per minute onto microfilm and electronic media simultaneously in one scanning pass. Bell & Howell's Filmer/Scanner is both a high-speed document scanner and microfilm camera that creates both an electronic and microfilm image of a document in one pass.

With optical disk technology, document imaging systems move businesses and government agencies closer to the paperless office by providing the means to digitally store paper documents in massive image databases that can be accessed in real time. Storing documents in digital form can solve many of the problems commonly experienced with paper files, such as misfilings that cause confusion and waste time, outdated documents that are mixed in with current documents, and paper that becomes damaged with age.

Having information stored in digital form can eliminate extensive, exhaustive searches for documents and files that tend to "disappear" in traditional storage facilities. Access to an image document can be effected by a variety of methods using content- and concept-based retrieval methods. Documents in either electronic or microfilm form can even be indexed in multiple ways that allow them to be retrieved in accordance with each user's informational needs.

With imaging systems in place, companies can realize other benefits that come with streamlining business operations. In the process of redesigning workflow to accommodate imaging systems, it is not uncommon for companies to find that they can combine several departments into one. For example, firms that previously employed separate work groups to process and store orders, invoices, and canceled checks are finding that the introduction of imaging systems enables all the documents to be stored on the same system. The queuing capabilities of the imaging system direct and monitor the flow of image files from sta-



tion to station on the network. Because the documents can be accessed from any workstation simply by cross-training employees to handle multiple job functions, several operations can easily be merged. This, in turn, means that fewer people are needed to handle the same functions. In the bargain, overtime pay can become a relic of the past.

This brings up another benefit of imaging systems: they can solve the familiar problem of multiple access. In a paper-filing system, only one person can access a document at a time. If that person brings that document back to his or her desk or—worse—takes it home, others are denied access until it is returned and properly filed. Prioritization schemes that attempt to determine who gets documents and when almost never work, causing frustration among staff members whose work is delayed. With imaging systems, documents can be called up simultaneously by multiple users. Some imaging systems even allow the documents to be updated by different users, whereupon the updated document is filed in the order of the last update first. That way, a complete history of a document can be maintained.

When imaging applications are supported by private networks or public network services, the benefits are even more pronounced. According to the Yankee Group, a market research company in Boston, network support of imaging applications can reduce staff needs by as much as 33% and can free up floor space by as much as 80%.

## 1.2 OFFICE DOCUMENT IMAGING SYSTEMS

There are four categories of office document imaging systems, each based on document size and other information-related factors. The size of the document is important because it impacts the type of input and output peripherals required, including the type of workstations needed, as well as basic architectural considerations.

### 1.2.1 General Business Systems

General business systems vary depending on whether they are oriented toward transaction processing or a central repository function. In a transaction processing environment, image management systems improve the economics of managing the active life cycle of the document. This type of system is used to automate workflow. In a central repository, the objective is to store vast quantities of images over a prolonged period of time. This type of system is used to archive documents after they reach a certain age, perhaps in compliance with various legal and regulatory requirements.

### 1.2.2 Image Processing Systems

Item processing is a special form of imaging in which small documents, such as checks and credit card receipts, are scanned in high volumes. Banks and other financial services firms scan millions of such documents daily. In 1990, the U.S. banking industry alone processed 50 billion checks at an average cost of only \$0.05. The image management sys-

tems in use today improve processing throughput by as much as 40% over previous generation systems and 80% over manual methods.

### 1.2.3 Engineering Drawing Systems

Engineering drawings typically depict complex, two-dimensional product specifications or design plans. Some of these drawings are 100 feet or more in length, depending on the level of detail required. They require the use of special-width scanners and plotters, in addition to specialized drawing software. Although many engineering firms employ *computer-aided design* (CAD) and *computer-aided manufacturing* (CAM) systems to create and manipulate drawings, these systems are not able to convert the massive base of paper drawings that have already been accumulated throughout the organization. An image system can bridge the two environments by electronically converting the paper drawings and storing them in image databases for easy access.

### 1.2.4 Mapping Systems

Mapping systems are typically used to manage land use by government agencies. They can be used to assess property taxes because of their ability to depict the amount of available land, the number of buildings on the land, and the number of new buildings the land will allow. Mapping systems can also depict any number of other features about the land, including the location of gas and water mains, underground wiring and pipelines, dump sites, and watershed and wildlife areas—anything that might be impacted by construction. These mapping systems rely extensively on color and graphics capabilities, as well as compatibility with the dissimilar systems used by other government agencies. Many of these maps are in paper form and could benefit from image management systems. When the maps are stored as images, they are often referred to as *geobases*.

There are a number of other types of imaging systems, such as medical imaging systems, holographic imaging, fractal imaging, and virtual reality. These specialized concepts are beyond the scope of this book, but are summarized in Appendix A.

## 1.3 IMAGING FUNCTIONS

The functions provided by document imaging systems differ according to vendor. There are several pivotal functions performed by imaging systems that have broad significance for potential users in any industry. These functions include:

- Scanning and indexing—scanning transforms paper documents into digital form for storage on magnetic or optical media and indexing documents allows for easy identification;
- Storage and retrieval, which are the basic functions of any imaging system;
- Compression, which applies an encoding algorithm to images for minimizing storage requirements;



- Workflow automation, which enhances basic imaging systems by facilitating control over various work processes;
- Office automation, which integrates imaging into a much broader information processing context.

### 1.3.1 Scanning and Indexing

*Scanning* is the process of reading paper documents into digital form. *Image scanning* produces a digital replica of the document, which can be displayed on a computer screen. This type of scanning reproduces corporate logos, drawings, photographs, and signatures with a high degree of accuracy. Editing tools allow the scanned items to be appropriately sized and cropped. Another type of scanning is *optical character recognition* (OCR), which is used to read machine-printed text and store it in word processor or *American National Standard Code for Information Interchange* (ASCII) format. Once in digital form, the text can be indexed, edited, written to floppy disk, or transmitted over the network.

### 1.3.2 Storage and Retrieval

Images are usually stored on high-capacity optical disks. The storage capacity of the common *compact disk read-only memory* (CD-ROM), is equivalent to approximately 700 conventional floppy disks. Improvements in lasers will double capacity in the next few years. *Write-once read-many* (WORM) optical disks and *compact disk recordable* (CD-R) optical disks provide more archiving flexibility and are becoming more affordable. Access time is also improving. Whereas 340 ms is typical for CD-ROMs, the access time for rewritable optical disks is about 80 ms. In a few years, 40-ms access time will be the norm. Another advantage of optical disks is the ease with which they can be integrated into a network via servers.

Retrieval is a basic function provided by all document imaging systems, differing in the speed of image retrieval and the amount of flexibility afforded to the user. Some systems allow the user to change fields and set up unique processes for accessing images. Others set up these processes in advance, making subsequent alteration difficult. Some vendors are oriented more to storage than retrieval, reflecting their backgrounds in paper and film records. These types of vendors typically serve customers with acute record management needs.

Several document retrieval approaches are available, including index and full text. With text-based retrieval, the user can access images using any word, rather than having to choose from a limited selection of key words in an index. With the index approach, the user enters a key word and is presented with a list of image files that contain that word. The user may have to call up several image documents before finding the item of interest. The benefits of the full-text approach are obvious: users can locate image documents based on their precise informational needs and inquiry preferences. Some imaging systems support both document retrieval methods and allow a fair degree of user customization.

### 1.3.3 Compression

Compression algorithms have become highly sophisticated, allowing image files to be shrunk by a ratio of at least 2 to 1 and as high as 100 to 1 without any loss of image quality. There are compression techniques that offer compression ratios up to 250 to 1, but with a slight loss of image quality.

Compression algorithms work by removing extraneous or repetitive data from an image. This permits more efficient storage, as well as faster transmission over *local-area networks* (LANs) and *wide-area networks* (WANs). When the document is retrieved from storage, it is decompressed within seconds so it can be viewed on a computer screen.

### 1.3.4 Workflow Automation

Beyond storing and retrieving images, workflow software automates the flow of documents from one processing step to another, eliminating many intervening processing stages and streamlining others. Workflow automation can be compared to an assembly line—workers in various departments or work groups collect information from many different sources and in many different formats and append the information to a core document until eventually a finished product is produced. A core document can, for example, be a purchase order from a customer that arrives by mail in paper form. In the mail room, the document is scanned into digital form and stored in a database for access by a customer service representative. The service representative files the purchase order in the customer's electronic file folder, which is ready for access by several other department workstations in sequence. The process ends when the order is shipped and an invoice sent via mail or *electronic data interchange* (EDI). The whole process is accomplished within a few hours rather than days or weeks, as in paper-based systems.

### 1.3.5 Office Automation

Rather than view document imaging as a separate application, businesses are coming to understand how imaging can be an extension of office automation. This entails treating images as just another type of data that can be combined with text, tables, graphics, and even voice and video. While integrating document imaging into the broader context of office automation is complicated, it offers significant efficiencies. Among other things, such integration can eliminate the need for redundant information system components—one set for images and another for other kinds of data. Although full integration of imaging and office automation is not yet common, it is expected to become widespread by the turn of the century.

## 1.4 IMAGING SYSTEM COMPONENTS

A complete imaging system typically includes the following components:

- Workstations with high-resolution displays;
- Scanners and OCR software, for digitizing hard-copy information;
- Software designed to manage document retrieval, workflow automation, and communications functions;
- Output devices, such as printers and plotters;
- Storage media—most notably the optical disks.

These imaging system components are illustrated in Figure 1.1.

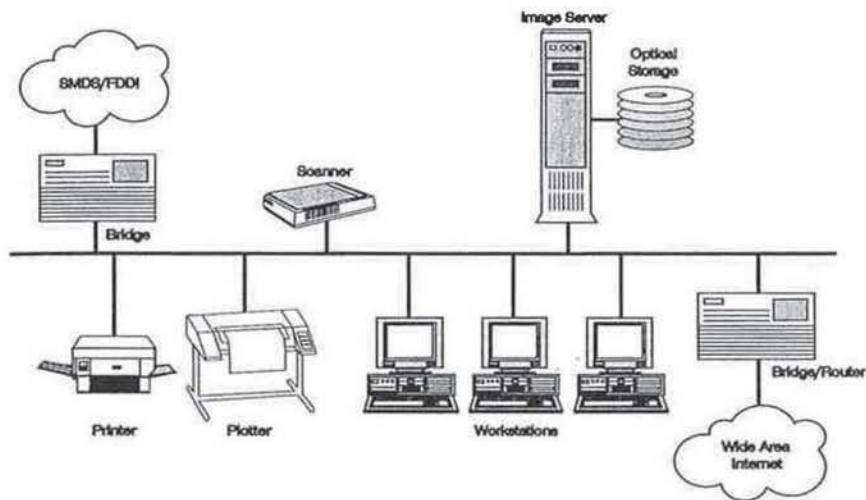


Figure 1.1 Document imaging system components.

Document imaging systems are available for microcomputer, minicomputer, and mainframe platforms. Moreover, imaging products address a broad range of user configurations, from single users, to work groups large and small, to enterprisewide operations. Prices are as diverse as the configurations themselves, ranging from only a few thousand dollars for a stand-alone desktop system to \$2 million or more for enterprisewide systems.

## 1.5 MARKET PLAYERS

The involvement of the major computer manufacturers is speeding market acceptance of document imaging. IBM, DEC, Unisys, Wang Laboratories, and Hewlett-Packard are among the traditional computer giants that offer advanced document imaging systems and planning services. Some, like Unisys, even offer their customers comprehensive educa-



tional programs that provide a management-level overview of the key issues, component technologies, applications, functionality, and motivations involved in imaging.

Among the newer (post-1985) entrants into the imaging system market is FileNet Corp., which is posing a serious challenge to the larger computer manufacturers. FileNet offers an array of high-end imaging systems and sophisticated workflow software. According to industry analysts, FileNet has more installations (but less revenue) than IBM.

While entire imaging systems that can support complex workflow applications and large numbers of users are available from big and small vendors alike, there are many more smaller firms that provide the discrete components of imaging systems, such as workstations, CD-ROM units and optical jukeboxes, scanners, software, and network facilities. These components can be put together by systems integrators to meet specific user requirements. The systems integrator is especially useful when there is the need to customize hardware and software interfaces to make components from different manufacturers work together in seamless fashion.

And, because the *regional Bell holding companies* (RBHCs) are free to enter the market for information services, analysts expect that these companies will eventually offer document imaging services to their business customers on a service bureau basis. This is a logical extension of the voice mail, electronic mail, and facsimile service bureaus that long-distance carriers, local telephone companies, and *value-added network* (VAN) providers now operate.

## 1.6 COMPLEMENTARY TECHNOLOGIES

A document imaging system can coexist with and complement other related technologies, such as EDI, multimedia, videoconferencing, and facsimile, further extending its benefits throughout the organization.

### 1.6.1 Electronic Data Interchange

EDI entails the transmission of machine-readable information between business partners' computers in a mutually agreed upon format. For example, instead of processing a purchase order with multiple paper forms and mailing it to the supplier, the data is passed through an application link, where software maps the data into the standard machine-readable data format. That data is then transmitted to the supplier's computer, where it is passed to an application link that maps the data into the internal format expected by the supplier's order entry system.

This interchange of data offers numerous benefits to users, including streamlined dataflow, which makes it easier to develop and maintain complete audit trails for all transactions. Having all of this information on line also provides the means to track vendor performance, perform cost-benefit analyses, improve project management, and enhance overall financial control.

EDI works well in established customer and supplier relationships in which both parties have approximately the same level of automation and adhere to agreed-upon proce-

dures for transaction processing. The problem with EDI is that the great majority of job functions either cannot benefit from this type of communication or require more advanced capabilities than EDI can provide. In these situations, document imaging systems are more appropriate. This is not to say that there is no point of convergence between document imaging and EDI. In fact, documents in image form can complement the documents sent via EDI. For example, although EDI is routinely used to deliver a purchase order for a part, it typically does not include a diagram of the part. Imaging can change that, allowing users to incorporate graphical information into the EDI document.

EDI is most often used in parallel with a document imaging system to process and manage invoices electronically. EDI facilitates the electronic exchange of documents between trading partners, and the imaging system's workflow software creates the link to accounting and imaging applications. In this way, the actions of coding an invoice, which is now an image, and updating the accounting package with the data can be combined into a single step. Some companies that use this approach have reported productivity gains of up to 25% and paper reductions of up to 40%.

### 1.6.2 Multimedia

The term *multimedia* generally refers to the integration of text, image, audio, and video within a single application program or online session. This integration of formerly separate applications has already found practical uses in education and business. In a training application program, for example, a CD-ROM contains all the information necessary to teach an installer how to set up and operate a new product for a customer. Through a combination of text, audio, image, and full-motion video, the installer can read the relevant technical documentation, view a demonstration of the setup procedures, and listen to a warning about possible electrical hazards—all accessed from the same program. When the installer completes a module of instruction, a built-in test facility can determine the installer's level of proficiency and provide suggestions for review.

Although there is a convergence between document imaging and multimedia—even multimedia that relies extensively on imaged documents—each is appropriate for specific applications. Multimedia is ideally suited for training and education, while document imaging is appropriate for document storage and retrieval, as well as for automating workflow in transaction processing environments.

### 1.6.3 Videoconferencing

When imaging is coupled with videoconferencing, documents can be viewed by all participants. *Videoconferencing* is the capability to broadcast audio and full-motion, slow-scan, and freeze-frame images over a closed circuit to one or more locations. Image, text, and graphics from a variety of sources can even be combined over a single video circuit, eliminating the need (and cost) of additional communications links to handle each traffic type separately. With a split-screen capability, an imaged document can be displayed on one side of the screen (or window), while the participants continue their discussion on the



other side of the screen. Participants who want a copy of the displayed document can request a fax or use screen-capture software to capture the image directly into their computers.

#### **1.6.4 Facsimile**

In a typical application of facsimile, a computer-generated text document is delivered electronically for output on paper. When imaging is added, graphics can be integrated with text for transmission. An incoming fax can be delivered directly into a document imaging system, skipping the paper process. More elaborate imaging systems allow documents to be faxed, or form letters printed, from within the imaging application itself.

### **1.7 NETWORK REQUIREMENTS**

Networked imaging systems are capable of more advanced applications than stand-alone systems. Stand-alone systems are little more than electronic file cabinets, whereas networked systems allow multiple access and facilitate information flow throughout the organization.

In the workflow environment, managers can distribute work among their staff. In turn, staff members use the network to access the information they need to fulfill their functions. In the end, a complete information package is produced. In the case of a home loan mortgage, a bank's imaging system can facilitate the title search, locate the property appraisal, find a photo of the house, call up the applicant's credit report, and process the loan application. This process, which once took many days or weeks to complete, can now be done in a matter of hours.

To fully realize the efficiencies and economies of imaging systems would require that documents be accessible not only to local users over the LAN, but also to remote users over the WAN. Imaging requires higher transmission speed and throughput, consistent data access, high-speed workstation interfaces, and interoperable communications protocols.

Most companies are still making the transition from mainframe-centered networks to local-area networks. In the former, access to host resources is typically provided through low- to medium-speed (4.8 Kbps to 56 Kbps) lines. Because imaging systems can impose a tremendous burden on existing networks by the sheer size of their files, such companies are finding it necessary to redesign their networks to handle the increased load. While a single-page text file might consist of 3 KB and a technical drawing one or more megabytes, the scanning process can easily double or triple the file size.

With many users simultaneously exchanging image documents, even 10-Mbps Ethernet and 16-Mbps token-ring LANs can quickly become overloaded. For Ethernet, associated network management overhead, access contention, and the possibility of collisions can reduce data throughput by more than 50% under heavy loads. A scanned paper document of 50 to 100 KB, for example, can take one second to transmit over a fully loaded Ethernet, while medical images or photographs of 2 to 5 MB can take 45 sec or more to

transmit. Token-ring networks offer a substantial increase in speed at 16 Mbps, but actual network throughput is often substantially lower because of network overhead.

Although the situation can be improved somewhat by using compression and decompression algorithms and by making the packet sizes as large as possible via the network drivers, a thorough review of the network topology might be required. The new topology would be determined by such factors as the frequency of image transfers and the size of the files. Other factors that bear consideration in the network modeling phase include average and peak input and retrieval volumes, average and maximum numbers of "pages," the retention period, the annual growth rate of electronic files, the number of shifts per day, and the number of individuals and productive hours per shift.

For imaging applications in the campus environment, LANs based on the 100-Mbps *fiber distributed data interface* (FDDI) standard can be used, especially for bandwidth-intensive medical imaging applications. Bridges and routers are available that connect Ethernet and token-ring LANs to FDDI backbones. Bridges more than routers, however, introduce an element of network filtering and forwarding delay that must be considered when planning such internets.

To support volume image traffic over a wide area, along with other voice and data applications, the backbone portion of the enterprise network requires at least T3 facilities, which provide a throughput rate of 44.736 Mbps over fiber-optic lines. But T3 equipment is proprietary and requires that the same type of equipment be used from end to end, even through the various carrier-serving offices that support the link. In essence, T3 is a custom service that entails high installation costs and recurring monthly costs.

The new carrier-provided metropolitan-area network services, such as *switched multimegabit data service* (SMDS) provide an alternative to expensive, private T3 fiber facilities and offer T3 as well as T1 transmission speeds. SMDS is a fast packet service that is optimized for high-speed data and image traffic. SMDS can even be used in conjunction with FDDI, extending the geographical range of the FDDI backbone across a metropolitan area.

Some local-exchange and alternative-access carriers offer higher-speed fiber services based on *synchronous optical network* (SONET) standards. Under SONET, these networks provide transmission speeds of 155 Mbps and 622 Mbps, with upgrades to gigabyte speeds planned for the future.

Other options for long-distance image transmission include universally available circuit-switched digital services, such as T1. For limited image traffic, there are also Fractional T1 circuits (64 Kbps to 768 Kbps) and switched 56-Kbps digital services. Newer offerings, such as AT&T's 384-Kbps and 1.536-Mbps switched digital services, can also be used to support certain types of image applications.

*Integrated services digital network* (ISDN) can also support image traffic. It offers some very tangible benefits, including faster call setup and network response times, bandwidth control, reconfiguration flexibility, and restoral options. Another benefit of ISDN is that the user pays only for the bandwidth used. If the organization's image traffic is infrequent and of low volume, ISDN bandwidth can be dialed up when needed. The user is



billed for the amount of bandwidth used (up to 1.536 Mbps), based on time and distance, just like an ordinary phone call.

Satellite services can also be used for image transmission between international locations. For example, images stored on optical disks in the United States can be retrieved by corporate sites in Europe or the Pacific Rim.

Whatever transmission facilities or services are used, provisions must be made to keep the wide-area links up and running. After all, companies are conveying business-critical documents, which means that imaging systems, as well as the networks they run on, must be robust in terms of reliability and availability. This means having redundant subsystems in place, spare bandwidth capacity available, and management tools accessible that can identify and isolate problems before they degrade overall system or network performance.

### 1.8 SELECTION CRITERIA

Selecting a document imaging system requires balancing the features and functionality with such information systems criteria as compatibility, architecture, and growth potential. Buyers need to find a system that will accommodate both initial and future applications.

Buyers should not limit their investigation of imaging systems and components to only the advertised capabilities, however. Of equal importance is the vendor's financial stability and ability to service and support the system at all locations. A careful evaluation will determine the vendor's ability to support products and systems well into the future.

During the evaluation process, it is important to look for a vendor who is willing to get involved with the end users of the imaging system, perhaps in a pilot project. This is an opportunity to take advantage of the vendor's expertise in on-the-job training, as well as an opportunity to spread the benefits of imaging throughout the rest of the organization and win over some skeptics in the process.

Other challenges await *information systems* (IS) and LAN managers planning image information systems, such as deciding whether a client/server or a host-based system is best and choosing the right peripherals.

### 1.9 IMPACT OF OPEN SYSTEMS

Open systems is a key issue in the imaging industry. Open systems offer a number of advantages, including transportability of applications across hardware platforms, flexibility in hardware selection, reduced training and support costs, and interconnectivity with external constituents, such as customers and trading partners.

Open systems interfaces begin with an operating system interface and extend to interfaces for graphics services, programming tools, languages, utilities, libraries, windowing and user interfaces, data and file management, and network facilities and services. Strict adherence to these interfaces allows different applications to communicate and exchange data across different devices, operating systems, and networks.



Several organizations are defining interface specifications and open system architectures, while vendors are attempting to develop products that adhere to preliminary specifications or that can be easily upgraded to the final specifications once they become accepted as standards. The characteristics of an open image system include:

- The availability of images in standard formats;
- The ability to exchange index data and images with other systems;
- Workstations that use standard user interfaces and run standard application software;
- The attachment of peripherals through standard interfaces.

Lack of standards in key areas of document imaging could impede its acceptance among companies that are concerned about such matters as premature system obsolescence and the offered system's compatibility with other vendors' products. Ultimately, users will drive the movement toward standards by choosing systems that offer the broadest range of compatibility. Users should try to select the most open imaging system possible, to support mission-critical applications, and should avoid wholly proprietary systems unless the offered features are technically essential. Since manufacturers have an interest in creating an environment that will promote rapid market growth—so they can recoup their research and development investments as quickly as possible—this point is not likely to escape their notice.

On the bright side, standards for open image systems already exist in many areas, including communications and networking, peripheral interfaces, user interfaces, workstation operating systems, database management system access, image format, and image compression.

Although open systems offer the highest degree of flexibility and investment protection, insisting on standards—while competitors are not—can result in a competitive disadvantage. This is because competitors will be in a position to reap the rewards of their imaging systems sooner, resulting in a lower cost of doing business. With the resulting savings passed on to their customers, they are in a better position to enlarge market share.

The standards process is unavoidably slow, since many diverse opinions must be considered and reconciled into a set of specifications that will be broadly accepted. Companies do not always have the luxury of waiting for standards before deciding which technologies to deploy. After all, they must be competitive in the world today to ensure their survival in the future.

### 1.10 DETERMINING NEEDS

If an organization's work environment is characterized by one or more of the following circumstances, it may be a prime candidate for a document imaging system:

- High volume of paper documents;
- High frequency of document retrieval;
- Complex, procedural orientation;

- Large clerical staff;
- Long document processing periods before data is finally made available to users;
- Corporate-level interest in improving customer service.

Implementing an enterprisewide document imaging system is not an easy task. Once the need for a system is determined, interdepartmental cooperation and coordination is essential to implementation success.

Successful implementation often involves breaking apart established business processes and recombining them in new and streamlined ways. The fact that many businesses do not adopt document imaging systems after extensive workflow studies and planning sessions is attributable more to the failure to resolve managerial issues than failure to overcome technical and logistical obstacles. This is understandable, since it is inevitable that the imaging system will have to be merged with existing information systems to provide enterprisewide benefits. This will necessarily have implications that cross departmental boundaries. Sometimes the prevailing belief among corporate managers is that their business processes have evolved a certain way for good reasons, and should not be changed—at least not yet.

It is likely that imaging will result in the elimination of some clerical jobs, especially if these jobs grew out of the need to administer paper records. However, staff elimination should not enter into the decision to move to a document imaging system. Many times, the clerical jobs can be redefined, allowing individuals to expand their responsibilities or, with some retraining, move into entirely new areas. In fact, adopting a document imaging system virtually ensures that few, if any, new paper-intensive administrative jobs ever have to be added to the corporate payroll.

At the same time, it must be emphasized that the introduction of imaging will bring with it new tasks and responsibilities. There may be the need to add staff to handle document entry and conversion, database indexing, system training, and on-site support. Clerical staff who are no longer needed to administer paper files can be retrained to fill some of these new slots.

Of course, imaging will force the restructuring of work within most organizations, resulting in changes in workflow, job content, and work habits in general. Imaging might even dictate the consolidation of work groups or whole departments to achieve optimal work performance. It will certainly eliminate redundant operations and result in less duplication of files. Speedy access to comprehensive databases promises improvements in staff productivity and customer service, making the effort at restructuring internal workflow entirely worthwhile.

### 1.11 CONCLUSION

Although imaging is still a rapidly evolving technology, it is already becoming indispensable at many companies where paper storage has reached critical mass. Imaging promises to impact businesses in a number of ways, not only reducing paper load but also paper-

work. In the process, business operations will become streamlined, resulting in new efficiencies and economies that will impact competitive position.

With the price-to-performance ratio continually improving on all components of document imaging systems—particularly for scanners, optical storage systems and applications software—imaging will not only become more attractive to more companies, it will stretch to embrace more territory, both in terms of actual distance covered and areas of use. As imaging increases in importance, there will be new opportunities to enrich current image applications and spur the creation of new ones.

Imaging systems also offer opportunities for entirely new product and service offerings. Businesses have barely begun to discover the myriad competitive opportunities image systems can present. Technological advances, primarily in mass storage, graphic display, and networking, have already cleared the way for full-fledged electronic image processing. No longer must daily mail delivery be a traumatic experience for today's information-dependent office.



## Chapter 3

### *Imaging Applications*

#### 3.1 INTRODUCTION

Imaging technology is rapidly changing the way companies do business. When coupled with workflow software, an imaging system constitutes a business strategy that can reduce costs and improve customer service, while ensuring the quality and timeliness of organizational response. These improvements are achieved by zeroing in on the principal cause of lost productivity: paper. It is paper that forces people into unproductive pursuits that hamper organizational performance and growth. Paper must be sorted, filed, retrieved, routed, and filed again by hand. This process is repetitive, tedious, time-consuming, and demoralizing.

Not only does this process have no end, but more paper is continually absorbed into the maelstrom, making it ever more difficult for the organization to extricate itself when pressed by market forces or economic conditions to improve performance. More importantly, such activities tend to sap the energy and creativity of employees, whose time and talents can be better used elsewhere.

#### 3.2 APPLICATION OVERVIEWS

The relevance of imaging is not limited to any specific industry. Hospitals, for example, are turning to document imaging systems to better organize files and reduce retrieval time, especially during emergencies, when fast access to patient data is essential. Along with eliminating file bottlenecks that can jeopardize patient care, imaging systems allow hospitals to trim administrative staff, thereby eliminating unnecessary overhead that tends to inflate the cost of medical services.

In the pharmaceutical industry there are a variety of applications that can benefit from document imaging systems. Clinical data submitted by physicians and research institutions must be processed quickly to discover trends in drug treatment tests. Organizing this massive data is critical to the marketing success of new drugs, since approval by the *Food and Drug Administration* (FDA) can be delayed if the supporting documentation is



incomplete or inaccurate. Not only can unnecessary delays cause the company to lose millions of dollars, they can also prolong the release of important new drugs to patients who might be in desperate need.

Telephone companies are using document imaging systems to improve customer service by enabling representatives to retrieve documents instantly via workstations in response to customer inquiries. Images of purchase orders, shipping statements, invoices, and other administrative documents are retrieved virtually instantaneously from *optical jukeboxes*, which are capable of holding as many as several hundred optical disks. Instead of losing up to 30% of documents through data entry errors, companies can hold document losses to 2% or less with a document imaging system.

Airlines, too, are making good use of document imaging systems. Maintenance operations alone can generate over 30,000 purchase and repair orders per year, each requiring ten or more supporting documents. Most of these documents must be stored for several years, as required by law. This can quickly add up to millions of documents that must be stored and retrieved. The maintenance manuals and parts catalogs for the Boeing 727 aircraft, for example, total 50,000 pages. Even when the catalogs are stored on microfiche systems, highly paid maintenance crews must take time out to search for the right document, a process that is neither efficient nor economical.

Government agencies and regulatory bodies rely heavily on paper documents. When they are not requiring paper documents from businesses, they are generating mounds of their own. Among the top federal imaging applications are records management and retrieval, electronic publishing, archiving, and applications associated with engineering drawings. Government demand for imaging is diverse but stems largely from the need to provide faster access to information at a reasonable cost. Among the types of information collected by state government agencies and municipalities on a continuing basis are: birth, death and marriage certificates; maps; deeds; environmental impact reports; building permits; police and court records; voter registration information; and corporate filings. An imaging system that delivers such frequently requested information to the public in 15 minutes, rather than two or three hours, is representative of such a need.

State motor vehicle departments will one day rely heavily on document imaging systems for issuing driver's licenses. On California licenses, for example, a color photograph, fingerprints, a signature, and text information are combined with a security hologram of the state emblem and a data-packed magnetic stripe containing such information as date of birth, license plate number, and license expiration date. All this information is stored on an image database and is available on line to law enforcement and government agencies, which together generate 20,000 to 30,000 requests a month.

Under the old microfilm record system, searches were quite labor-intensive and, therefore, expensive. Another problem was the poor quality of the two-inch-square black-and-white photographs. By comparison, the new image system produces crisp color images that can be viewed on screen or output to a color laser printer in less than a minute. Already, spin-off applications are being developed that make use of the magnetic stripe on the new licenses. Police officers will one day fill out tickets simply by passing the magnetic stripes of the licenses through a hand-held device that will issue a printout to the traf-

fic violator. The hand-held device will also download accumulated citations into the traffic ticket system, eliminating the need for paper records.

There are other reasons why businesses of all types and sizes are turning to imaging systems for document storage and retrieval. Records managers are responsible for ensuring that documents are kept long enough to satisfy the legal requirements of various government agencies. Failure to produce relevant documentation at the required time can result in huge fines or even criminal charges—if it is found that records were carelessly or deliberately destroyed.

Satisfied with the integrity of document imaging systems, government agencies are now more willing to accept documents in electronic form. For example, on December 31, 1991 the *Internal Revenue Service* (IRS) adopted Revenue Procedure 91-59. Accordingly, the agency no longer requires companies to store paper documents, as long as their electronic equivalents provide the same level of detail as the originals. There are other conditions attached to electronic storage. Documents must be kept as long as their contents might be relevant to tax laws, records created on older systems must be retrievable by replacement systems, and documentation—including software changes and security controls—must be maintained on the electronic filing system. And should IRS examiners show up, the company must provide all computer resources necessary for them to audit the electronic books.

Actually, it was inevitable that the IRS would come to accept the validity of imaged documents, since for years it had been accepting—even encouraging—the filing of tax forms in electronic form. To file electronically the user sends a file via disk or modem to a third party, called an *electronic return originator* (ERO), usually specified by the tax program. The paper forms are also sent, separately. These forms include the U.S. Individual Income Tax Declaration for Electronic Filing (form 8453, included in the software) and the W-2 forms. The ERO sends back a receipt, which serves as a bill to the IRS if taxes are owed, and sends the IRS the return. If the user is owed a tax refund, the IRS sends a check or makes an electronic deposit into the user's checking account, usually in a matter of a few weeks. ERO charges for this service are between \$20 and \$30.

In addition to allowing users to file tax forms electronically and enabling the IRS to process the information and issue refunds faster, tax preparation software offers users several other advantages. Some programs interview users to determine filing status and the forms that must be used. Some help the users get organized by providing a series of prompts and fields that must be filled in via keyboard entry. The programs plug each item into the correct form and schedule and then total the entries. An audit feature alerts users to incorrect or missing data, as well as excessive deductions. Modules for filing state taxes electronically are usually available separately.

The tax programs not only guide users through the maze of forms, but also assist them in figuring out what deductions to claim. Many of these programs include online tutorials, documentation, and context-sensitive help. In addition to suggesting possible deductions, some programs offer what-if analyses and provide financial planning and investment advice for the next tax year, taking into account such factors as inflation and cost-of-living increases.



Businesses, government organizations, and tax payers are not the only groups who stand to benefit from document imaging. Perhaps the most information-intensive organizations are libraries and universities. They, too, are abandoning microfilm and microfiche systems and moving to optical disk-based imaging systems. Some college dormitory rooms are wired not only for phone service, but also for a LAN connection for plug-in computer access to the university library and to other research facilities where data and imaged documents can be accessed from optical disk storage.

Public libraries are also implementing optical disk storage systems. Records of books and their locations are distributed via optical disk to all libraries belonging to the state system. Optical disks are also starting to replace the traditional card catalog for listing books by title and author. Some libraries are preserving historical documents by capturing them in image form for storage on optical disk. Researchers, instead of poring over old originals—and, in the process, inadvertently contributing to their destruction—can obtain the same information by scanning *imagebases*. And they can find what they want much faster than with microfilm and microfiche.

Eventually, financially-strapped universities and public libraries will awaken to the possibilities of using document imaging technology to offer value-added research services to such niche markets as the scientific, medical, law, and engineering professions. For example, a lawyer representing a client in a malpractice suit might need background information on a particular medical procedure. From a modem-equipped computer, the lawyer could dial up the library's optical disk server and initiate a key-word search to obtain the required information. The information could be in the form of images of medical journal pages or papers delivered by experts at various symposia. If the local library did not have the requested item, the request could be passed on to a library that did. The library could not only charge an access fee for the imagebase, but also charge for printing out hard copy from a particular imagebase—with a portion of the printing charge going to the copyright holder.

State and local government agencies could also increase their revenues by offering value-added services. An architect looking for a specific building design that meets stringent state codes for earthquake tolerance, could, for example, access the Building Department's image files for the blueprints. By seeing how another architectural firm met the state requirements, the architect could get an idea of what was actually involved before submitting planning and cost information to a client.

This chapter explores the following real-world applications of document imaging and workflow software: technical documentation, credit card operations, accounts payable transactions, loan processing, and insurance claims processing. From these examples the reader can gain insight into how document imaging and workflow software can be applied to other applications.

### 3.3 TECHNICAL DOCUMENTATION APPLICATION

Most companies develop and issue technical documentation to acquaint users with how to assemble, install, operate, and maintain their products. The amount of technical documen-

tation varies from company to company, depending on the sophistication, number, and diversity of products offered. Many companies also document their own manufacturing operations, processing operations, and quality control procedures.

The types of information that comprise the technical documentation might be quite diverse and could include internally generated drawings, photographs, customer specifications, parts lists, schematics, blueprints, memos and correspondence, and vendor manuals. The sources of these items might also be diverse; the information could come from customers, suppliers, subcontractors, strategic partners, government agencies, and standards organizations.

### 3.3.1 The Problems

Typically, different groups in an organization will make use of its documentation; however, the groups will customarily use different parts of the information, depending on what is relevant to their job duties and responsibilities. Some of these groups include purchasing, manufacturing, quality assurance, maintenance, engineering, customer service, document control, and training. Each group is usually aware of the existence of other, related information—it is just a matter of tracking the information down, if and when it is needed. It would be unthinkable to provide every group with its own copy of all documents, even if that were possible. So, when one group needs information held by another group, a lot of time is wasted in pursuing it.

Logistical problems aside, it would not be very practical to distribute documentation to all groups within an organization. The reason is that the information typically exists in many different forms. Much of it might exist electronically in the form of word processing files, CAD/CAM drawings, or compound documents that contain text, graphics, and images. Information distributed in these forms requires specialized workstations and software. And because these types of documents typically do not use a common interchangeable format, multiple workstations and software would be required—an expensive proposition for most companies.

Even if a company could afford to equip each group with all the specialized hardware and software needed to view multiple document types, the problem becomes one of change management. It would be necessary to have procedures in place to ensure that any changes made to one copy of a document were carried over to all other copies of the document, which might be scattered around the world via the enterprisewide network. Furthermore, it would be necessary to have in place a mechanism for approving change requests, so that all parties impacted by the change could be made aware that a change was being considered and have a chance to add their own input to the change process. Depending on the procedures in place (Figure 3.1), the whole approval cycle could take weeks or months. And there is always the possibility of documents getting delayed or lost. To implement such procedures would require a document control group, which often makes the problem worse. In a competitive environment, the whole process of changing and approving documents can result in lost business opportunities.



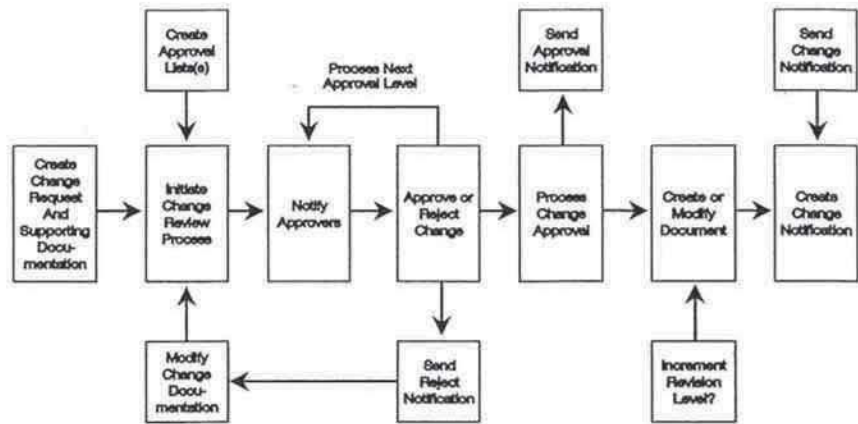


Figure 3.1 The typical document-change control process. Source: ACCESS Corp.

### 3.3.2 The Solutions

Document imaging systems bring together a number of technologies that allow users to capture, store, manage, retrieve, display, distribute, and print information that originates on paper or film or is generated by computers.

The imaging process starts with a scanning of the document, which is similar to photocopying. Then each scanned image is indexed by using automatic character recognition or by reading the punched Hollerith data on an aperture card.<sup>1</sup> The scanned images are stored in digital form on magnetic or optical disks. Computer-generated files, such as CAD/CAM files, can also be stored on the system. Often these files are stored in both native file format and raster format, which allows viewing on either the original CAD system or any workstation capable of viewing scanned images.

A number of different types of scanners are available to handle documents of different paper sizes and weights, documents that are bound or double-sided, large engineering drawings, photographs, aperture cards, jacketed microfiche, and film.<sup>2</sup> Some scanners can handle a duty cycle of tens of thousands of pages daily. (By way of comparison, Hewlett-Packard's market-leading LaserJet IIIp printer has a maximum duty cycle of only 8,000 pages per month.)

<sup>1</sup>Aperture cards combine the attributes of punched cards and microfilm. The punched card contains a description of the document and acts as a frame for holding a strip of microfilm. The cards can be processed to yield full-size paper copies of the microfilm image.

<sup>2</sup>A variety of media are used for document images, including opaque, translucent, linen, vellum, clear, or matte-finish film. Some scanners can accommodate all of these media types, in sizes of up to 35 in wide.

The disks are housed in optical storage and retrieval libraries, or *optical jukeboxes*, consisting of multiple 5.25-in or 12-in optical disks. The stored documents can be retrieved in a matter of seconds for display on high resolution workstations or PCs. Multiple jukeboxes can be connected together, allowing an organization to expand its storage facilities incrementally as needs dictate.

Once in digital form, the documents can be managed more effectively and at a single point. Documents are indexed—manually or automatically—by the imaging system's software at the time of scanning. The index information is written to a relational database. When users request a document, they can be sure of getting the most up-to-date copy. Distribution employs digital networks and standard telecommunications facilities, which also allow multiple users to view the same document simultaneously. When T1 lines are used, an E-size engineering drawing in raster form can be sent within ten seconds of the request. Over fiber facilities, the same document can be sent in a fraction of that time.

Because raster images are used, it is possible for most types of workstations and PCs sold today to be used for document display—even low-end displays costing only a few hundred dollars—particularly when such displays are used only for checking the quality and alignment of scanned images.<sup>3</sup> Thus, expensive high-resolution displays and CAD/CAM software are not necessary to view such documents. Because of the wide availability and low cost of the software needed to view raster images, these images are the most practical form for the widespread electronic distribution of documents. An authorized user can make changes at a CAD/CAM-enabled workstation and work with a file in its native format. Others can overlay annotations at image-enabled PCs.

Once scanned, oversized documents of C-size and larger are stored in tiled raster image format. In this format, each image is saved as a set of logically connected square tiles, invisible to the user upon display. This is done automatically, without operator involvement. Tiling large documents provides more efficient and faster retrieval of the desired viewing area.

Each document stored as a tiled image has an associated preview image created at scan time. This preview image is a reduced image of the document that allows a user to see each sheet or page without having to pan from side to side or up and down if a total document exceeds the viewing area of the display. When retrieving a large document stored as a tiled image, it is the preview image that is seen first by the user. This permits the user to find the desired page for viewing without wasting time looking at each page in turn. The user selects the area to be viewed by using a mouse to point to the area of interest. The tiles for this area are then retrieved and displayed at full size, without any visual indication of the tile boundaries.

The user can toggle between the preview image and the full-size image, selecting the desired viewing area interactively. The user can browse between documents or pages within a document, move up and down or side to side, and use the zoom feature to view

<sup>3</sup>A good argument can be made for providing more expensive, high-quality displays for operators who work with images all day long as part of workflow processing applications. Better displays would greatly reduce eye strain and fatigue.



details within a document. The images can be rotated by keyboard command or can be made to rotate automatically upon display, if that feature is selected during scanning. Users can select from rotation values of 90°, 180°, and 270°.

The imaging systems software includes capabilities that allow users to perform redlining (Figure 3.2), markup, and editing on the document images. Each user's modifications to the image document occupy a separate layer, which can be distinguished by color. The various layers can be saved and managed along with the underlying document and later merged with the underlying raster image. The comments, sketches, and geometric shapes entered by users in the various layers can be cut and pasted to the image, area-shaded, or block-erased. Users can modify the image by selecting edit symbols from a raster library. Users who are not authorized to initiate document changes would only be allowed to review the documents and would be locked out of the raster symbols library.

Storing documentation as image files can save a lot of time and money spent in finding, accessing, viewing, and editing individual items—time and money that would normally be wasted if the documents were available only in paper form. And, for organizations involved in preparing technical proposals to government agencies and commercial vendors, document imaging systems can greatly reduce the time spent in finding and consolidating information from diverse sources and locations.

### **3.4 CREDIT CARD APPLICATION**

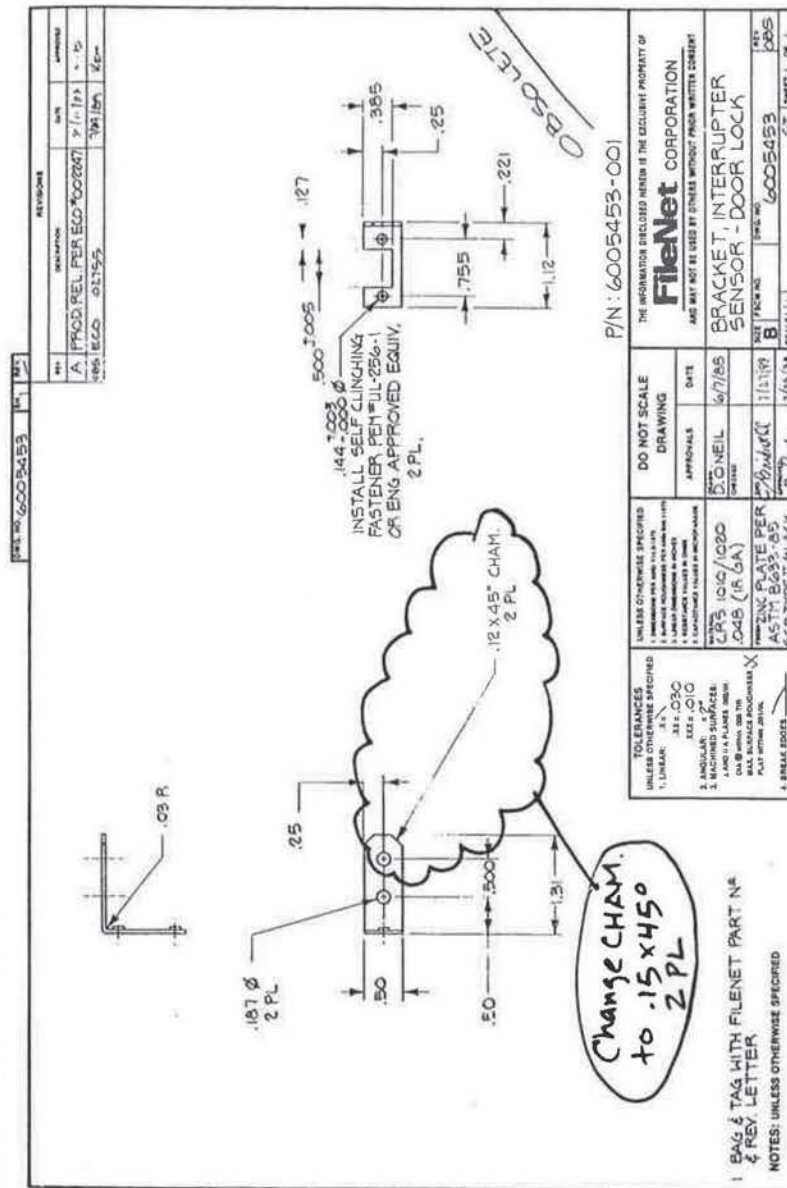
In today's troubled economy businesses of all types and sizes are wrestling with ways to stay afloat without jeopardizing their competitive positions over the long term. The banking industry is no exception. To survive in the recession, banks too are streamlining operations and cutting back staff. Along with dealing with the vagaries of the recession, banks are being compelled to offer more and better services to retain existing customers and attract new ones. Not only are banks competing with each other for a larger share of the credit card market (Figure 3.3), but nonbanking institutions are also entering the market and competing against banks.

#### **3.4.1 The Problems**

Credit cards are one of the vehicles banks are using to improve their competitive positions. In the span of a decade credit cards have replaced the checking account as the banking industry's most profitable product, responsible for 30 to 60% of an average institution's total profits. Credit card operations are also among the most costly services to maintain, because of huge amounts of money lost on bad credit, fraud, and delayed resolution of customer billing disputes. State and federal laws ensure fair and timely reporting of credit information and protect consumers against inaccurate and unfair billing practices.

Moreover, credit card issuers must resolve inquiries and disputes in a timely, efficient manner. All of this has forced banks to take steps to reduce the financial risks associated with this type of product and to increase the level of service, to maintain strong revenues from their credit card programs. This is hard to do because credit card operations





**Figure 3.2** Sample redlined drawing. *Note:* The revision to the original image occupies a separate layer, which can be distinguished by color and can be erased and changed again without permanently altering the original. *Source:* FileNet Corp.

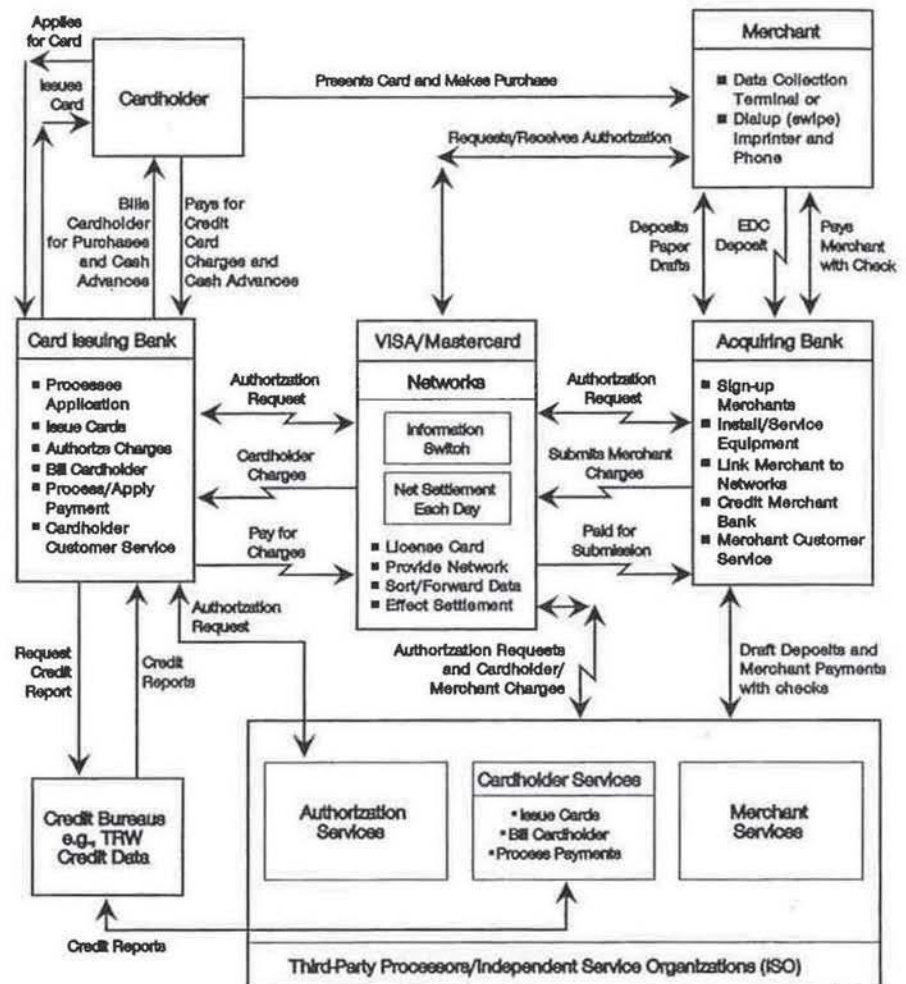


Figure 3.3 Credit card processing. Source: FileNet Corp.

are highly paper-intensive. On the other hand, this type of operation is a prime candidate for integrated image processing solutions.

Mainframe computer systems are capable of automating only approximately 10% of the information involved in credit card transactions. The other 90%—chargebacks, balance requests, address changes, finance charge adjustments, payment disputes, and card applications—all involve continuous streams of incoming paper.

Imaging systems that make use of optical disk technology, high-resolution workstations and PCs, and powerful process-automation software can work with an organization's existing mainframe programs to further automate the numerous paper-intensive tasks involved in credit card operations. An imaging system can bring together the different types of information for online access from a single workstation. That way, customer service representatives can resolve inquiries during initial calls, without having to track down the required information and call customers back.

### 3.4.2 The Solutions

A typical credit card transaction begins with a bank customer submitting an application for credit. Based on the available information and the use of various scoring and reporting systems, the bank decides to approve or deny a credit card to the applicant. The less information available, the more risk involved in the decision process, both in approving the wrong applicant and denying the right one.

An imaging system can speed up the entire application processing cycle. Since all applications are scanned into the system, the need for data entry is greatly reduced, eliminating keying errors. Manual data entry can be confined to taking phoned-in applications. Once scanned, all applications are immediately available for online access and can be automatically routed to the appropriate representatives for processing. Representatives can track application status from their desk on an initial call, rather than having to call the customer back days or weeks later.

Fraud, which can cause extensive financial loss, is a major concern of most credit card issuers. Imaging systems can reduce fraud risks by allowing banks and other institutions to increase the quality of approvals. When an imaging system is connected to a mainframe, decision makers can receive the most up-to-date information on each applicant, including online fraud files, customer histories, and third-party scoring and credit reporting data—all information that might be located on different hosts. Users can access all of this information—and update it instantly—from an imaging system workstation.

Organizations can also use the imaging system to enhance customer service. Most credit card issuers receive thousands of inquiries daily from cardholders and merchants who honor the cards.<sup>4</sup> Billing disputes, balance requests, address changes, finance-charge adjustments, requests for new cards, and statement and insurance questions are among the types of calls that banks receive. These inquiries generate callbacks, outgoing correspondence, customer file and history updates, and billing adjustments. In addition, credit card issuers must honor requests for original sales slips by either keeping the paper sales slips, keeping a microfilm version of slips on file, or requesting copies from merchants.

To retain customers and gain new ones in a competitive market, inquiry resolution must be fast and the results satisfactory. However, paper files and microfilm tend to inhibit quick inquiry resolution.

<sup>4</sup>A large bank like Chemical Bank in New York—the nation's sixth largest—receives 60,000 inquiries by mail and phone each month. The bank uses an imaging system from FileNet Corp.



When a cardholder disputes a charge on a monthly statement, the credit card issuer temporarily credits the cardholder's account until the dispute is resolved. If the merchant bank has proof of charges, the merchant represents the charge amount back to the issuing bank. A merchant might resubmit a transaction deposit up to three times for payment and, if not satisfied, might request a final arbitration hearing. The dispute is resolved when either the amount is permanently credited to the cardholder's account or the cardholder pays the amount to the issuing bank to be credited to the merchant account.

This process is complex, and there are strict time limits for processing chargebacks and guidelines for submitting supporting documentation if errors are claimed by either side. If these time requirements and guidelines are not met, the issuing bank might be forced to credit the cardholder, regardless of who is at fault.

Imaging allows multiple users immediate workstation access to all files (Figure 3.4), resulting in the resolution of most problems during the initial call. Imaging also enables

Figure 3.4 shows a computerized document imaging system interface. The left pane displays a scanned document of a letter from F.C.R.A. dated November 28, 1988, with handwritten initials 'C/B' and 'd'. The top right pane shows a mainframe window with cardholder account history, including fields for name, address, phone, and various account numbers. The bottom right pane shows a 'CUSTOMER SERVICE INQUIRY SYSTEM' menu with options like '1. FICA/Good Customer Bailiwick (1022A)', '2. FC Collection on Pur/Cash Rev. (1021A)', etc.

Figure 3.4 Multiple views of customer information. *Note:* Incoming mail is scanned, indexed, and stored. When a customer calls with a problem, a customer service representative can quickly resolve the problem with an automatic display of customer information in multiple windows: card-holder correspondence (left side), a mainframe window with card-holder account history (top right), and a control window that guides the operator through the inquiry resolution steps (bottom right). *Source:* FileNet Corp.

credit card issuers to meet processing time limitations on disputes, easily handle the numerous types of outgoing correspondence to customers, update customer and fraud files in real time, and eliminate expensive microfilm and paper storage facilities.

An imaging system can be used to store customer signatures and geographic data, thereby helping users track customers and confirm proof of legal debt responsibility. In the past when a credit card account defaulted, the bank would search for the original signed application in paper or microfilm files, then make paper copies for distribution to the collections group. When an unresolved collection problem went to court, lawyers required the original applicant signatures as evidence. At some banks half the time the original documents could not be located for presentation in court, resulting in lost revenue from increased chargeoffs. With image processing lawyers always have the collection documentation they need when they go to court.

Image processing can help prioritize work, reducing the organization's exposure to financial loss from the untimely resolution of chargebacks. Each stage in the work process can be assigned a specific time frame for completion. If a document does not progress past a particular stage within its assigned time frame, it can be brought to the attention of the appropriate customer service representative or automatically routed to a supervisor.

In addition to controlling the movement and processing of work, an imaging system automates a number of other management tasks and can provide a significant reduction in the time needed to handle inquiries and resolve disputes within legal-charge time limits. For example, an imaging system can monitor the number of outstanding disputes and track the progress, age, and resolution of disputes through a number of productivity reports. These reports allow supervisors to adjust employee workloads and schedules to ensure timely processing of all work.

The workflow scripts that make all this possible can be developed by the customer with assistance from the imaging system vendor. Simple inquiry operations can be fully automated within 30 days, while more complicated operations, such as chargeback processing, can be automated in two or three months. Other operations, like fraud analysis and security, can be phased in later.

One leading vendor of imaging systems, FileNet Corp., claims to have helped banks dramatically improve a variety of operations with its FileNet imaging solutions. According to the company, FileNet provided these improvements:

- Increased application approvals by 50% or more;
- Resolved chargebacks in half the time;
- Doubled the number of resolved inquiries;
- Reduced fraud cases by as much as 45%;
- Reduced required training time by 50%;
- Eliminated microfilm resource requirements and maintenance costs;
- Reduced data entry by 50%;
- Improved document retrieval by 99%;
- Improved general productivity by an average of 40%;
- Provided savings to clients of from 30 to 60% per item processed.

An added benefit of imaging systems is increased job satisfaction for customer service representatives—they are freed from the daily burden of paper files and dissatisfied customers. Instead of looking for paper files, they spend more time helping customers, providing them with personalized attention and service and enhancing corporate image in the process. The lower stress levels also result in a decrease in employee turnover, saving the company hiring and retraining costs. Not only can fewer staff handle more work, but, when hiring does become necessary, the time it takes to train new customer service representatives is greatly reduced. With an imaging system in place, a trainee can become productive in a week instead of months.

### 3.5 ACCOUNTS PAYABLE APPLICATION

Organizations tend to think of their accounts payable operations as quite different from those of other organizations, despite the similarities in transaction types and accepted accounting principles. The common denominator, of course, is paper.

#### 3.5.1 The Problems

Each day invoices arrive in the mail and must be matched with original purchase orders and delivery receipts. A Fortune 500 company can pay one million bills or more every year for goods and services and can handle more than 80,000 invoices per month. An accounts payable workload of this magnitude requires automation.

Although financial application programs run on mainframe systems make purchase order information available on line and eliminate some of the paper in accounts payable departments, a substantial number of invoices still remain on paper. The management of that paper continues to consume corporate resources and delay timely payments to vendors. Even without exceptions, it is not uncommon for vendors to wait three months or more for payment, with two months being the average wait.

In addition to coping with paperwork, some common problems accounts payable departments face include:

- Incorrect or duplicate payments;
- Invoices that get lost or left unpaid;
- Failure to prioritize invoice processing to meet vendor payment terms and take advantage of associated discounts;
- Failure to match invoices with supporting documents;
- Failure to answer vendor inquiries in a timely manner.

#### 3.5.2 The Solutions

Although a mainframe can handle 90% of the transactions in an accounts payable department, it contributes little to resolving the remaining 10% of transactions, called "exceptions." These are the transactions that involve out-of-file, misfiled, or lost paperwork.



Clerks can spend an inordinate amount of resources finding, routing, and finally processing the paper originals.

Typically, the accounts payable process starts with the generation of a purchase requisition for goods or services. The purchasing department then reviews the available vendors, selects a vendor, and generates a purchase order, specifying the items required, purchase prices, and shipping and billing information. When the goods are shipped, the vendor mails the customer an invoice requesting payment. The amount of goods received might be entered into the mainframe inventory control database, or the paper delivery receipts might be filed in a storage cabinet, to be retrieved later when payment is requested.

The accounts payable department, usually a centralized accounting function, receives the invoice and is responsible for matching at least three items: proofs for goods or services purchased, received, and billed.

Most organizations have clearly defined policies and procedures to guide personnel through much of this paper-intensive workload (Figure 3.5). Handling exceptions, however, can cause significant delays in vendor payment because of the paperwork required and the number of people who need to get involved in resolving the problem. Some examples of exception cases include overshipment of goods, incorrect shipments, price discrepancies, overbilling, and duplicate invoices.

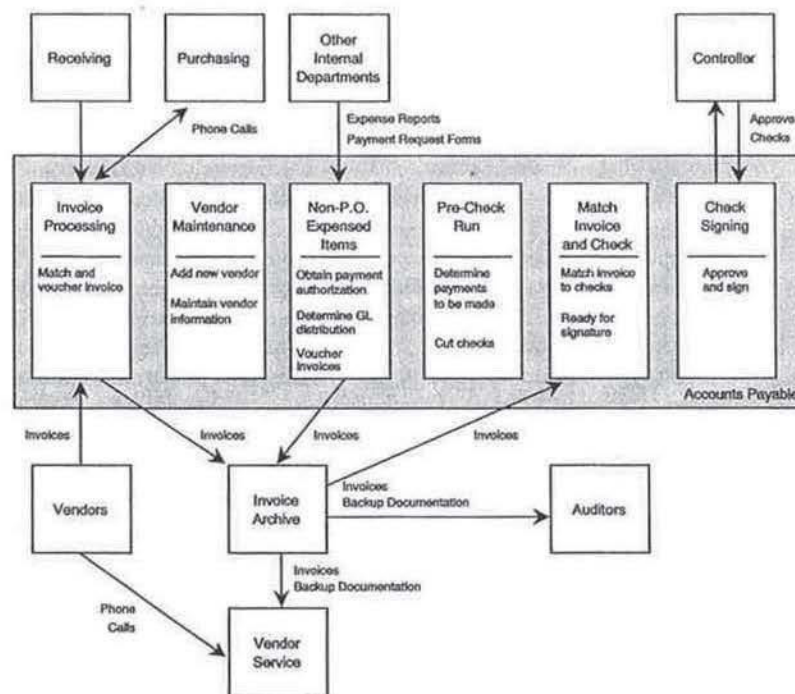


Figure 3.5 Typical accounts payable process. Source: FileNet Corp.

In addition to processing invoices from vendors, the accounts payable department is also responsible for processing such payouts as employee expense reports, rents, utilities, taxes, and other nonpurchase-related expenses.

Procedures for controlling, monitoring, prioritizing, and handling exception cases can be automated with an imaging system. Invoices can be automatically matched with delivery receipts. With immediate access to invoices and backup documentation, accounts payable personnel can provide better service to vendors on initial phone calls, without leaving their work areas. The following scenario illustrates how an exception case is resolved with the aid of an imaging system.

Invoices entering the accounts payable department are batch-scanned and date- and time-stamped (Figure 3.6). The scanned invoices are then indexed with vendor codes and routed to the appropriate accounts payable clerk. The clerk reviews the invoice and enters invoice data into the mainframe accounts payable system. In this case, internal checks show a difference between the billed amount and the purchase order amount. Further investigation reveals an item price discrepancy. The clerk calls the buyer for resolution and puts the invoice "on hold" for one week, using the imaging system's holding queue. The clerk then moves on to the next invoice in the electronic in-basket.

Workflow software manages the storage, retrieval, processing, display, communication, printing, and routing of all information required to process work. After one week, the imaging system's workflow software automatically selects the problem invoice and routes it to the supervisor at a high priority. The invoice moves to the head of the supervisor's in-basket for processing. The accounts payable clerk is free to work on other activities, and the invoice is not lost or forgotten.

Many times purchase orders, contracts, invoices, and receiving documents must be matched up to assure that materials or services have indeed been received. The imaging system's workflow software provides a "rendezvous queue" that holds scanned documents until receipt of other appropriate supporting documents. The workflow software tracks outstanding items and notifies the accounts payable clerk when all documents have arrived or when the document has been on hold beyond its pre-assigned time limit.

In cases in which the various documents are scattered among several mainframes, the imaging software's terminal emulation capability allows clerks to look up vendor information in one or more of these systems to complete invoice processing. Because the imaging system can be integrated easily with mainframe accounts payable applications—often without requiring any changes to the mainframe programs—clerks can access inventory and customer data, as well as document images of invoices, through multiple windows opened at the same workstation. This feature allows clerks to compare documents from various sources without leaving their workstations.

The workflow software also includes a variety of system-generated reports that allow supervisors to monitor the productivity of each accounts payable clerk. This permits the reallocation of work among employees to balance their workloads. In addition, the imaging system can produce an audit trail of documentation flow, which aids dispute resolution with vendors.

[illegible]

Figure 3.6 Data entry screen. Source: FileNet Corp.



An added benefit of imaging systems is that they can retrieve documents for auditing purposes without disrupting the daily operations of the accounts payable department.

### **3.6 INSURANCE PROCESSING APPLICATION**

Underwriting, which is one of the most critical functions in the insurance business, is the interface between the insurance company and the customer. It is performed by either an independent agent or a policy holder. It is the principal means by which insurance companies gain new revenues and minimize risk.

#### **3.6.1 The Problems**

The underwriting process usually requires an inordinate amount of clerical and professional resources. Depending on the insurance product, client files can grow to include hundreds of documents. The files can be physically passed dozens of times between various professional and clerical personnel before a single application is fully processed. Paper files are often suspended in tubs while awaiting the arrival of other required documents and are not readily available to representatives handling customer requests. In addition to consuming a lot of space, even the best managed paper-based system makes the information in each file available to only one person at a time.

Claims processing is another important activity of insurance companies, and it usually requires ongoing interaction with customers. Customer satisfaction is determined by the insurance company's speed and efficiency in processing claims, as well as by its ability to accurately respond to questions about customer status. The insurance company's ability to perform in these areas determines whether customers stay with the insurer or transfer their business elsewhere. If, for example, the insurance company administers the employee health care plan for a Fortune 500 company, employee dissatisfaction with the claims process could result in the loss of a sizable key account.

Prompt customer service can be difficult to achieve if paper files are tied up in the claims department. Such delays could result in duplicate files being created, further extending work backlogs and possibly resulting in duplicate claim checks being issued.

For customers who are independent agents, the tasks of the insurance company are to track the agents' state licenses and credentials and keep the agents up-to-date on the various policy types and services. All this work can add up to a major paper trail, which must be managed to ensure timely service and new revenues. Some insurance companies have licensed and contracted 20,000 or more agents and do business with them on a regular basis.

Annuities represent another booming business for insurance companies. This area also produces an incredible stream of documents, including: account and loan applications; payment amount changes; withdrawal requests; account balance requests; fund transfers; name, address, or beneficiary changes; and numerous other documents associated with pension plans. Literally millions of sheets of paper can be accumulated pertain-

ing to "live" executive pension plans. In addition, every shift in the economy produces a flood of phone calls from investors requesting changes in their annuity plans.

### 3.6.2 The Solutions

Voluminous files are digitized for permanent storage on high-capacity optical disks, where they can never be lost or misfiled and are accessible to multiple users simultaneously (Figure 3.7). Workflow software can be programmed by the user to emulate and improve established paper-processing flows.

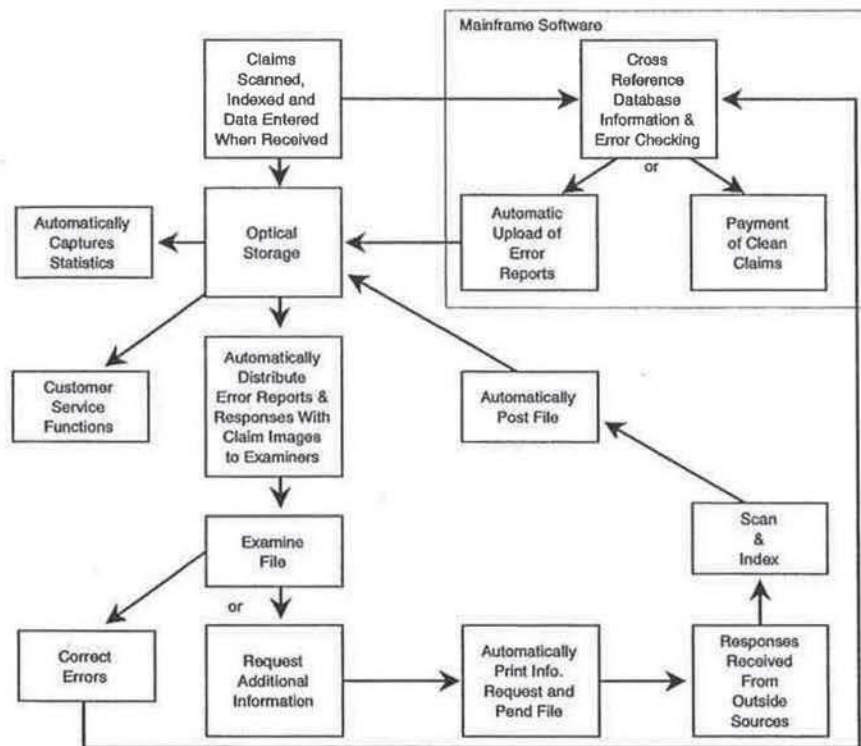


Figure 3.7 Typical claims process as handled by an imaging system. Source: FileNet Corp.

Some workflow software includes features specifically designed for insurance companies, such as the ability to rapidly group together documents into folders. The underwriter can browse quickly through the folders, marking or tabbing documents as needed to note important items, even magnifying hard to read sections, and electronically paper-clipping notes to any document or file.

An imaging system allows users to view portions of documents within a file, access a mainframe computer for additional data, and generate and automatically print form letters for routine requests or customized letters for unusual requests. All these tasks can be accomplished without clerks having to leave their workstations.

Often insurance companies' information databases contain sensitive information that must be protected by law, such as HIV test results for AIDS. Imaging systems can provide multiple levels of password protection to control access to certain functions or files.

Workflow software can also manage the state licensing of independent agents, by keeping track of the receipt and status of all necessary documents and the distribution of documents to service representatives for processing. The software can generate custom reply letters, electronically annotate special conditions or circumstances in documents, and distribute correspondence from the agents. With all documents stored in digital format, files are available anytime the agent calls to request status or other information.

In the very competitive annuities market, insurance companies must struggle to maintain consistent profit margins and keep levels of customer service high—in the face of rising clerical costs for managing all the paperwork. An imaging system with automated workflow capabilities can facilitate processing of all the necessary documents, from initiation to final disbursement. The system can interface with the mainframe computer to track account balances, master indexes, and all the intervening documents. Complex transactions can be completed in parallel by several departments because each department has simultaneous access to the same information.

Some imaging systems also offer interfaces to facsimile, so transactions can be entered directly into the imaging system via fax transmission and confirmed via fax, without creating additional paper records.

### **3.7 LOAN SERVICES APPLICATION**

In the last few years the recession, increased competition, and a plethora of variable-rate loan products have combined to shrink profit margins in the loan servicing industry. Lenders are being challenged to access documents more quickly in answer to a barrage of questions from consumers regarding a wide variety of products. While striving to improve customer service, these institutions have also had to cope with escalating labor costs.

#### **3.7.1 The Problems**

Since economies of scale are important factors in determining loan servicing profitability, experts predict a trend toward greater industry consolidation in the years ahead. In the short term, however, lenders must attempt to service larger volumes of mortgage, home equity, consumer, and student loans. Many long ago automated their operations with computers. But this step only automated about 10% of the information needed to support their business operations. The remaining 90% of information originates outside the organizations and enters through the mail room. A huge volume of information must be sorted, routed, processed, filed, retrieved, and refiled in original paper form. The collected infor-



mation does not readily lend itself to traditional data processing; it is in the form of typed and handwritten documents, graphics, and photographs.

Over the life of a mortgage or home equity loan it is not uncommon for the loan file to quadruple in size. And, because an office can hold only a certain amount of paper files, the records management center is often at a different location than the department requiring the documents. Off-site record storage substantially increases file retrieval time.

To further complicate matters, loan servicing activities are controlled by the availability of documents in the loan folder files. A high percentage of loan documents are out-of-file when requests are made—another clerk has the file. Sometimes two clerks are processing the same document without knowing what each other is doing. Overall processing time to complete departmental transactions is delayed by the lack of an organized and controlled document routing system. As a result, multiple copies of the same document are made to satisfy the requirements of several departments simultaneously, which frequently causes more confusion and delay, not to mention problems associated with determining which copy is the most complete and updated.

The ramifications of such file-processing inefficiencies can be quite serious. Taxes and insurance might not be paid on time from escrow accounts, which might result in interest penalties and, in extreme cases, property liens. Also, prompt processing of insurance renewals and policies is needed to protect lenders from loss. If complete, error-free loan documents are not presented to investors in a timely manner, lenders could face interest penalties or cancellation.

### 3.7.2 The Solutions

Many lending institutions are turning to advanced imaging technology to cope with the avalanche of paper and to increase staff productivity and responsiveness.

Instead of having to locate, move, track, copy, box, ship, and refile paper-based loan files—often under extreme time constraints—the imaging system allows documents to be retrieved in seconds, so that responses to customer inquiries can be handled quickly, efficiently, and economically. Imaging also leads to improved management of the volumes of paper required to process and track tax and insurance payments in a timely manner, reducing the risk of penalties.

Lomas Mortgage USA, holder of one of the nation's largest mortgage loan portfolios, valued at \$20 billion, reports an improvement in document retrieval time from four and one-half days to 10 seconds, using an imaging system provided by FileNet Corp. The processing error rate for records management declined from 2.4% to 0.4%, while the unit cost per loan dropped from \$1.85 to \$1.39. With the streamlined workflow and efficiency provided by image processing, Lomas was able to reduce the staff headcount by half.

### 3.8 CONCLUSION

The potential of imaging technology to radically transform a business from a lethargic, marginally profitable operation to a thriving, highly profitable one has been amply demon-

strated over the years. Examples abound from virtually every sector of the economy of companies that have improved performance and cut costs to stay competitive in dynamic markets.

Improvements such as these are achieved by attacking one of the most persistent root causes of poor performance: paper. Dealing with ever increasing volumes of paper is counter-productive. It forces organizations to maintain artificially large staffs and to continually replace people who are lost to job stress or want to pursue more challenging careers elsewhere. Paper slows response time to customer inquiries, which not only encourages customers to do business with competitors, but also prevents organizations from successfully pursuing new business opportunities. If allowed to continue, the limitations imposed by paper hamper not only organizational performance, but long term growth as well.

Imaging systems can impose order on the chaos created by paper. In this sense, imaging is not just a technology, but a key business strategy that can enhance an organization's future competitiveness—to say nothing of its short-term survival.

## Chapter 6

### *Considerations for Mainframe-Based Imaging*

#### 6.1 INTRODUCTION

Since the 1960s mainframe computers have become a permanent fixture of most large corporations and government agencies, despite repeated predictions of their demise in favor of distributed computing and peer-to-peer networking via LANs. The sizable investment in software applications alone is enough to keep mainframes entrenched in many organizations for the foreseeable future.

According to the META Group, a research firm in Westport, Connecticut, corporations worldwide have invested \$750 billion on software applications for IBM or IBM-compatible mainframes. And, despite the recession, users spent \$26.6 billion in 1991 for 14,000 new mainframes. According to Dataquest International, this represents a drop of 6.4% from the previous year.

In addition to being the only platform many companies trust to run their mission-critical applications, there are several other reasons why companies are not quite ready to jettison their mainframes.

First, mainframes are capable of handling enormous transaction volumes—as much as 20,000 to 25,000 per hour—in addition to other applications, such as materials management, receivables, payables, payroll, and electronic mail. Adding document imaging to the existing host environment is relatively easy and permits many corporations to further leverage their investments in mainframe systems.

Second, because the mainframe offers a single repository, data is more readily consolidated, updated, and protected. Moreover, access to the data is more effectively controlled. The same levels of data integrity and access control in a distributed environment require resource-intensive software and continuous vigilance on the part of the LAN administrator.

Third, the mainframe can help an organization keep pace with user application needs. Although centralized host environments have long been criticized for application backlogs, this is becoming a nonissue with the continued evolution of host-based *computer-aided software engineering* (CASE) tools and *fourth-generation languages* (4GL).



The availability of such tools is also making the development of custom applications easier and faster. Whereas users of microcomputers and workstations tend to rely on off-the-shelf software that is immediately available and provides a common feature-function set, it is the custom applications developed for mainframes that often translate into competitive advantage.

Finally, over the years many organizations have developed a wealth of application code for their mainframes. Moving to a distributed systems architecture would mean re-writing or possibly throwing out major portions of that code. In most cases recompiling mainframe programs for a distributed environment is fraught with problems, making success difficult, if not impossible. On the other hand, it is a relatively simple matter to run third-generation application code through a machine filter to prepare it for migration to the 4GL environment.

The question that begs an answer is why sales of mainframes have experienced slow growth in recent years. There are three reasons. First, slow growth is the natural result of a maturing industry. With the majority of batch processing demand largely satisfied, future growth will come primarily from customer service and from vendors meeting the increasing demand for information processing and office automation applications, such as document imaging.

Second, the role of the mainframe is changing. It is evolving from a back office, number-crunching behemoth to a highly flexible, intelligent hub for corporate networks. In fact, the emphasis of vendors, including IBM and Unisys, is on improving the connectivity of their mainframes for enterprisewide computing. There are several components to this concept, including:

- Client/server computing, in which UNIX-based mainframes act as information hubs, serving the diverse needs of all departments;
- Mainframe channel extension over WAN facilities that support network-routing capabilities for resource sharing and disaster recovery;
- Interoperability between the host environments of different vendors, such as Unisys connectivity to IBM's SNA.

When this emerging role of mainframes is better understood and appreciated among users, mainframes will again experience healthy sales, although probably not on the scale witnessed in the early to mid-1980s. Already, however, mainframes are supporting document imaging applications, including those of Amdahl, Bull HN Information Systems, Hitachi Data Systems, IBM, Tandem Computers, and Unisys.

Third, the recession has affected the sales of mainframes in recent years, as it has for other types of computer and communications products. And, while fewer mainframes have been sold, they are also getting more powerful and less expensive, giving buyers more bang for the buck. According to a report by the research firm Computer Intelligence, the power of mainframes in terms of *millions of instructions per second* (MIPS) tripled between 1988 and 1991.

However, a more important measure of performance, from an imaging perspective, is total system bandwidth. This refers to two things: the amount of data that can move across the system's *backplane* (the computer's major internal communications path) and the total input/output bandwidth of the system for network and peripheral connections. When optical fiber is used for the backplane, internal throughput is measured in terms of gigabits per second, instead of megabits per second, which is used with conventional metallic backplanes.

A variety of fiber-optic channel extenders are available to support high-speed connectivity between the mainframe and remote peripherals and networks. Channel extenders based on the FDDI provide host-to-host or host-to-LAN connectivity at up to 100 Mbps. Channel extenders based on the *high-speed serial interface* (HSSI) provide the same type of connectivity at up to 52 Mbps. And channel extenders that support T3 offer connectivity between the hosts and remote peripherals at up to 44.736 Mbps over the WAN.

The T3 channel extender allows the mainframe to support multiple sites over high-speed lines, such as T1 and T3. When the host needs to communicate with a remote peripheral, it encapsulates the data in a packet, along with the target device's subchannel address, and routes it to the extender serving the remote site (Figure 6.1). In a full mesh network of channel extenders connected by T1 or T3 lines, switching channel extenders permits traffic to be automatically routed around failed communications links.

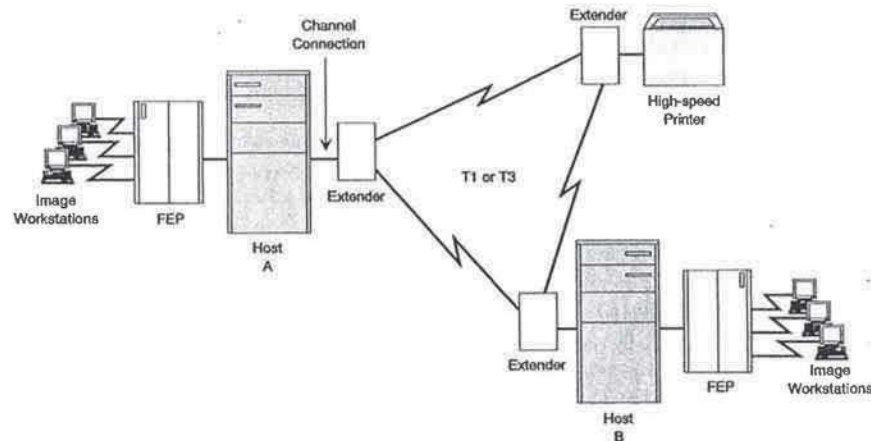


Figure 6.1 Wide-area channel networking.

The performance and networking capabilities of today's mainframes have vendors recasting their machines as LAN servers, performing such rudimentary services as file storage, print serving, and backup services.



## 6.2 ROLE OF THE MAINFRAME

In supporting high-volume document imaging applications, the mainframe essentially performs the functions of a file server and workflow manager. As a file server, the mainframe is functionally positioned at the center of the document imaging system; it coordinates the activities of the system's major components and manages vast amounts of imaged documents generated by the workstations. The mainframe receives new imaged documents from scanner-equipped workstations, distributes processing tasks to individual workstations, collects the results, and performs any necessary file consolidations. The mainframe also receives user requests for display, print, facsimile, and other processing functions.

As a workflow manager, the mainframe matches documents in need of processing with the appropriate or next available workstation operator. Routing is determined at the time the document imaging system is configured for operation and is based on such factors as user profiles, which describe the categories of work each user is trained to handle. Alternatively, the supervisor can create or change the user profiles to balance the work load. Either way, the workflow manager uses this information, along with information provided by indexing, to determine who processes what documents. To keep track of work in progress, the workflow manager maintains a processing history for each document, which is available for display at an operator's or supervisor's workstation.

In addition to functions for queuing and routing documents, workflow management can also include such functions as user- and document-access security, and support for backup and recovery procedures.

## 6.3 MAINFRAME IMAGE SYSTEM COMPONENTS

High-volume image processing and distribution in the mainframe environment involves the use of many types of equipment that are already used to support existing applications. In addition to the mainframe itself, these include terminals or workstations, printers, communications controllers and LAN or WAN connections, and magnetic and optical storage. Of course, the document imaging application must include a high-speed scanner, which is the one piece of equipment that usually needs to be added to a workstation.

### 6.3.1 Workstations

Most workstations will be used for viewing imaged documents. There are a variety of monitors available for this purpose. Color display monitors from 14 in to 20 in can display up to 256 colors simultaneously. These are useful for image applications that include artwork, color photographs, or CAD/CAM drawings. Otherwise, high-resolution gray-scale or monochrome monitors are adequate for most document imaging applications.

Monitors are available in portrait or landscape orientation. There are full-page portrait monitors that are vertically oriented to provide a maximum viewing area of 8.5 by 11.8 in. Portrait monitors are used when it is important that the user be able to view an entire page of an imaged document at one time, rather than having to scroll to view portions



of a page. Most monitors are horizontally oriented, or landscape. A virtual requirement for viewing engineering models, blueprints, and CAD/CAM drawings, a landscape monitor also provides a number of viewing options for the display of data and images (Figure 6.2).

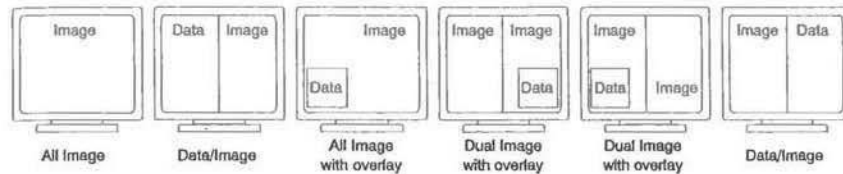


Figure 6.2 Data and image display options on landscape monitors. Source: IBM Corp.

Split-screen landscape monitors can display coded data and imaged documents side-by-side in separate display areas, allowing users to view host application and image sessions simultaneously.

In the IBM environment a display configuration called the *dual-headed configuration* has two monitors sharing the same workstation. Certain IBM monitors (models 8506 and 8508), when used for image display, do not permit the simultaneous display of host application coded data. This requires a second display monitor attached to the same workstation or a separate IBM 3270-series terminal (Figure 6.3).

### 6.3.2 Scanning Equipment

The scan workstation provides the means of capturing paper documents electronically. Capture can be performed either in a batch mode, in which all pages are continuously scanned with high-speed equipment, or interactively, with the pages being scanned and viewed individually. Interactive scanning allows users to display documents on the screen to check resolution and alignment and then make equipment adjustments prior to the high-speed batch scanning of similar-quality documents.

An optional component of the scanner is software that provides OCR. This software allows the scanner to read and extract typed data from the imaged documents—the entire page or particular zones—and convert it into ASCII text or a supported word processing format for editing at the workstation. Most standard type fonts are supported by the OCR software, including those printed in dot matrix format. OCR software flags characters it cannot recognize with a symbol such as the pound sign (#) so the operator can easily find them and make the corrections manually. OCR software can even be “trained” to recognize fonts or nonstandard symbols it does not normally support. This is done by scanning a training page containing such fonts and symbols and entering a string for the OCR program to use when it encounters them during scanning.

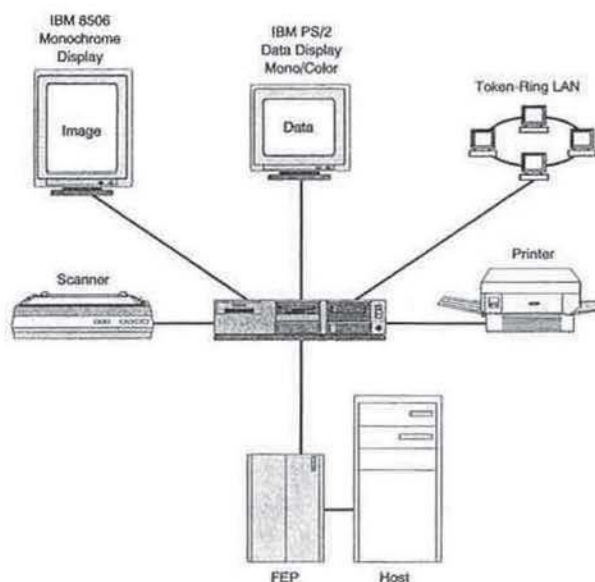


Figure 6.3 IBM's dual-headed monitor configuration. Source: IBM Corp.

Mainframe-based imaging systems excel at high-volume transaction processing. This starts with the rapid automatic capture and indexing of paper documents. The scanner can include the capability to read bar codes and imprint the sequence number on each image. Another advanced capability of high-speed scanners includes duplex scanning with single-pass transport. This allows two-sided pages to be scanned in a single pass. Without this capability, the page is flipped over and then rescanned. Either way, sides one and two are marked as such, and the images are collated by the system.

Output from the high-speed scanner can be used as input by the imaging system's folder application. This application uses a batch storage process to index and store documents and optionally route them for processing through workflow queues.

Other technologies can be combined with scanning to further automate document processing. Transaction-intensive environments, such as the banking industry, need very sophisticated systems for check processing. The discrete steps that must be performed include capturing the image; performing character recognition, endorsement, and encoding; and sorting the documents into separate pockets. The NCR 7780, for example, accomplishes this by performing high-speed *magnetic ink character recognition* (MICR) and OCR, enabling the system to interpret the code lines on a document. The NCR 7780 also uses *character amount recognition* (CAR) technology, which allows the system to scan, read, and store handwritten numbers on a document.

In addition, the NCR system performs "power encoding," which allows a large number of documents to be encoded at high speed. The NCR 7780 captures images, performs recognition and encoding, and endorses checks without the need for operator interaction. In the power encode mode, the NCR 7780 can process full payments as well as minimum payments at rates up to 10,000 transactions per hour.

### 6.3.3 Printers

Band printers, which are capable of operating at 70 lines per minute and more and use fan-folded paper, are common in mainframe data processing environments; but most imaging applications require the high resolution of a laser printer. A variety of laser printers, desktop and system-oriented, are supported by mainframe-based imaging systems.

Since document imaging systems are used ostensibly to eliminate paper, when printing is necessary, a conventional desktop laser printer offering 300-dpi resolution should be adequate for most applications. For image applications that demand higher resolutions, such as commercial publishing, 600-dpi, 800-dpi, 1000-dpi, and 1200-dpi printers are available.

Lasermaster Technologies' Unity 1000, for example, is a multiplatform printer that provides 1,000-dpi printing in minicomputer and mainframe environments (and also supports Macintosh and IBM-compatible PCs). The printer includes an internal hard disk with 135 preinstalled typefaces and features an Ethernet expansion port, automatic port switching, and halftone enhancement technology, which improves the appearance of halftone images by providing increased gray levels at higher screen frequencies.

Such high-resolution printers typically cost two to 10 times that of a 300-dpi laser printer, which averages \$2,000. Printer economics will change significantly in 1993, however, when Hewlett-Packard releases its 600-dpi desktop laser printer. The price tag is expected to be under \$3,000.

Most laser printers handle both standard and legal-size sheets, although a few printers that are based on RISC architecture, from QMS and others, output documents measuring up to 11 by 17 in at 20 pages per minute.

If color is important, there are a variety of color printers, starting at \$5,000, that print up to seven colors by thermal transfer or ink on plain paper, with a resolution of 300 dpi and better.

While most desktop printers perform at four to eight pages per minute, a printer in the mainframe or LAN environment would have to perform at 20 to 30 pages per minute, possibly more. There are new laser printers that can output 100 pages per minute with a 2 million-page per month duty cycle. Such a printer, the 300-dpi HP-F100 Jumbo Jet, is available in Europe. Developed by Hewlett-Packard and marketed by Fareham, Hampshire-based European Business Group Ltd., the Jumbo Jet supports all major minicomputer and mainframe hosts.

Users can print imaged documents in a variety of ways; they can print a single page from a folder, the entire folder, or several folders in batch mode. Not only can individual pages be printed, document-related notes, annotations, and document forms within a page



can also be printed. Print commands are selected from a menu at the view workstation. When set by the supervisor, a print request banner identifying the job owner is printed with each job. The supervisor can modify the list of print jobs, release jobs from queue, and inquire as to the status of jobs. The supervisor can also prefetch documents during nonpeak hours for high-volume retrieval and printing requirements.

#### 6.3.4 Communications Controllers

The mainframe environment is no longer isolated from the rest of an organization's computer resources. In fact, an enterprise can greatly benefit when mainframe image applications can be extended via the wider connectivity offered by LANs and carrier-provided facilities. LAN support is a feature that is now available with most communications controllers or front-end processors, as is X.25 and, increasingly, frame relay.

*Communications controllers* are devices that sit in front of the mainframe and control access to a relatively limited number of ports. The communications controller can provide connections to remote communications controllers at other hosts via such carrier-provided services as X.25 and frame relay, or private facilities such as T1.

Via plug-in interfaces, communications controllers can provide access to both Ethernet and token-ring LANs. In addition to token-ring support, IBM's 3745 communication controllers, for example, support Ethernet LANs that run the *transmission control protocol/internet protocol* (TCP/IP). This enables Ethernet-attached workstations to access IBM host resources, such as database information and applications, and provides backbone support for connecting Ethernet LANs across an IBM WAN. The 3745 also provides access to frame-relay networks, which provide more efficient data transport than T1 alone.

Another type of communications controller found in the mainframe environment is the *front-end processor* (FEP), which provides more configuration flexibility. The front end is connected directly to the mainframe, while the back end provides connections to various communications channels. The controller transmits and receives messages, assembles and disassembles packets, and detects and corrects errors. In the LAN environment, these functions are provided by the network adapters. FEPs also perform polling, allowing a large number of devices to share a relatively small number of mainframe ports.

FEPs provide *systems network architecture* (SNA) connectivity over services like X.25 and frame relay via routers. Of the two services, frame relay is preferred for image traffic. The FEP passes the SNA traffic to the router, where it is encapsulated in the frame relay protocol. At the remote location another router performs the process in reverse.

#### 6.3.5 Storage Facilities

Although desktop, LAN, and mainframe imaging systems use essentially the same magnetic and optical media for storage, there are some storage configurations used more often in the mainframe environment that can effectively support document imaging applications. These include magnetic tape and optical libraries, as well as *direct-access storage devices* (DASD) and *redundant arrays of inexpensive disks* (RAID).

### 6.3.5.1 Library Facilities

*High-end library storage facilities* are complete installations potentially consisting of thousands of individual tape cartridges or optical platters, which are retrieved by robots.

IBM, for example, offers its 3496 Tape Library Dataserver, an automatic loading tape cartridge retrieval system. The Dataserver consists of an enclosed structure along which a robot runs. Up to 64 tape transports can be attached along one side of the enclosure. From 5,660 to 18,910 cartridges can be installed—and operate unattended—up to 14 mi from the host, through the use of fiber-optic channels available under Escon.

IBM's online optical disk subsystem, the 3995 Optical Library Dataserver, can store up to 181.6 GB of images and other information on erasable optical platters. This is a greatly expanded version of the optical jukebox used on LANs.

IBM is not the only vendor who offers library storage facilities. StorageTek offers products for robotic tape libraries. The company also offers *Expert Library Manager* (ExLM), a software product that automates movement of tape cartridges used in its 4400 and 9310 *automated cartridge systems* (ACS). Cartridges are exchanged at a rate of 175 and 350 cartridges per hour, respectively.

Another StorageTek product, *ExtendedStore*, is a library that archives tapes containing less frequently used data. The product does not include tape drives. *ExtendedStore*, which can handle about 6,000 tape cartridges, keeps the data available for easy access. When data from *ExtendedStore* is needed, it is moved via StorageTek's pass-through port system to a standard library, which has an attached tape drive from which the data can be read or written.

For companies with less voluminous storage needs, StorageTek offers *TimberWolf* and *WolfCreek*, smaller-capacity libraries that handle 500 and 1,000 cartridges respectively, with a cartridge exchange rate that starts at 350 per hour. Each product can be upgraded in increments of 500 cartridges and can be attached to any StorageTek library. The open architecture of the StorageTek products makes them compatible with numerous IBM and non-IBM vendor platforms.

For organizations that want to file their image documents separately from other corporate data, there are tape jukeboxes that accommodate 10 to 20 tape cartridges. When 8mm tape cartridges are used, a tape jukebox of 10 cartridges can store 50 GB of image files. These units allow users to automate unattended backups and recoveries.

Sheer storage capacity is an important difference between the disk and tape media typically used in the PC and mainframe (and LAN) environments. With regard to tape, several vendors offer the new 8mm magnetic tape systems, which provide up to 5 GB of storage capacity per cartridge. Such systems have a data transfer rate of 500 KB per second and a search rate of 37.5 MB per second, allowing any image file on a 5-GB cartridge to be located within 85 seconds.

These 8mm tape systems use the licensed technology developed by Exabyte, which is the lone supplier. The ISO recognizes Exabyte's 8mm data format—used in its EXB8200 8mm Cartridge Tape Subsystem—as a worldwide standard. When used with an optional data compression module that is based on IBM's *improved data recording capability* (IDRC), 5 to 1 compression is available, which the user can turn on and off. When



on, a single 8mm cartridge can have a capacity of 25 GB, making such systems viable alternatives to more expensive optical disk technologies.

Vendors are enhancing the file search capabilities of 8mm tape systems. Winchester Systems, for example, offers Flash8mm, which incorporates Touch Technologies' Dynamic Tape Accelerator to increase backup rates by allowing the tape to stream in one direction continuously, instead of back and forth, in the manner of reel-to-reel tape systems.

With price performance continually improving, 8mm tape systems are migrating from mainframe, midrange, and LAN environments down to the work group and individual user levels. There, they are replacing the long popular *quarter-inch cartridge* (QIC) tape systems.

#### 6.3.5.2 *Direct-Access Storage Device*

Another type of storage facility commonly found in mainframe and midrange environments is the *direct-access storage device* (DASD), originated by IBM. Although the market for DASD is dominated by IBM, there are several third-party DASD suppliers, including EMC, IPL, Memorex Telex, and StorageTek.

In essence, DASD provides consistently fast access to data, regardless of its location in massive databases. Typically, the available storage space of DASD is allocated to specific applications, so that frequently used information is made immediately available. Less frequently used information is stored on less expensive archival media. In an imaging application, for example, active documents and file folders would be stored on DASD for quick retrieval, while inactive files might be stored on magnetic tape. Files that are considered closed would be permanently stored on optical media, such as WORM disks.

Organizations that use DASD also most likely use IBM's *Database 2* (DB2), a relational database management system. (IBM recommends DB2 for use with its ImagePlus document imaging systems.) However, the problem with DB2 is that it has a large appetite for DASD space. Considering the storage requirements of image files, any needless use of DASD space can greatly inflate the cost of imaging operations. According to a 1991 survey by Candle Corp., DB2 wastes large amounts of DASD space at an annual cost of between \$40,000 and \$100,000 per organization. On very large systems the cost of wasted DASD space can run into millions of dollars.

Not coincidentally, Candle offers a utility program that generates a financial analysis of how much money the DB2-allocated space that goes unused costs to maintain per month and recommends ways to reclaim unused DASD space. This is particularly useful because IBM's DB2 provides no way to determine the amount of storage space that is over-allocated or unused. The utility, DB/DASD, also provides an automatic allocation feature, which allocates the available space more efficiently, freeing database administrators from the time-consuming, laborious task of manually reallocating DASD space when they do discover unused DASD.

IBM does offer *system-managed storage* (SMS) products to assist database administrators in managing DASD use, but they do not support DB2 files. This forces database administrators to use third-party DB2 DASD management solutions.



DB/DASD does not, however, compress the database or otherwise reduce the amount of storage space that DB2 takes up. For these tasks database administrators must turn to products like DB2 Reduce, from Boole & Babbage, and Shrink/DB2, from Sterling Software. These products compress DB2 tables during the read/write process from and to DASD so the tables take up less space. But the products do not address the problem of how to reduce the amount of unnecessary space allocated to DB2. Instead, they reduce the amount of space a particular DB2 application uses within the allocated DB2 data partition.

For example, an image database might reserve 40 MB of DASD space at the time is set up. The actual DB2 files might take up only 20 MB when the database is put into operation. With compression that 20 MB might be squeezed into 10 MB—but there will still be the original 40 MB of DASD reserved for DB2 that cannot be used for image documents and files. This means that the effective use of DASD requires both a management utility and a compression product.

For disruptions caused by internal problems, there are three recovery techniques commonly used with DASD: journaling, checksum, and mirroring. *Journaling* is a system function that provides an audit trail of activity on physical files. It allows users to save files faster by saving only changed records and provides recovery of individual files to the point of failure.

*Checksum* protection is a system function that can help avoid a time-consuming system reload, even when a failed disk drive must be replaced. The system is able to reconstruct the data on any failing drive on the checksum set after a single drive failure.

*Mirroring* provides automatic duplication of data on separate storage devices. This ensures that the system will continue to operate and the data will remain completely intact when one of the devices fails. This approach allows users to keep the system available during and after DASD or other storage hardware failure, reducing or eliminating downtime for repair or recovery. However, mirroring does not eliminate the need for backup or guarantee continuous, uninterrupted operation. Nor does it make journaling obsolete. In fact, mirroring uses twice as much DASD, which is already expensive.

Channel extenders can support remote DASD over a metropolitan area, via high-speed private lines or carrier-provided services, or in campus network facilities, via such interfaces as FDDI, HSSI, or T3. These arrangements allow DASD to reside in remote locations as far away as 8 to 10 km, which, in turn, permits DASD to be used for electronic vaulting. In *electronic vaulting* a duplicate image of local DASD is maintained at a remote site to act as backup system.

#### 6.3.5.3 Redundant Arrays of Inexpensive Disks

*Redundant arrays of inexpensive disks* (RAID) is an emerging storage technology designed to address the data reliability deficiencies of DASD. In essence, a RAID configuration consists of multiple small disks. Instead of risking loss from a crash that could wipe out all of the image documents and files on one high-capacity disk, an organization can use RAID to distribute the image files and documents across multiple smaller disks (Figure 6.4).

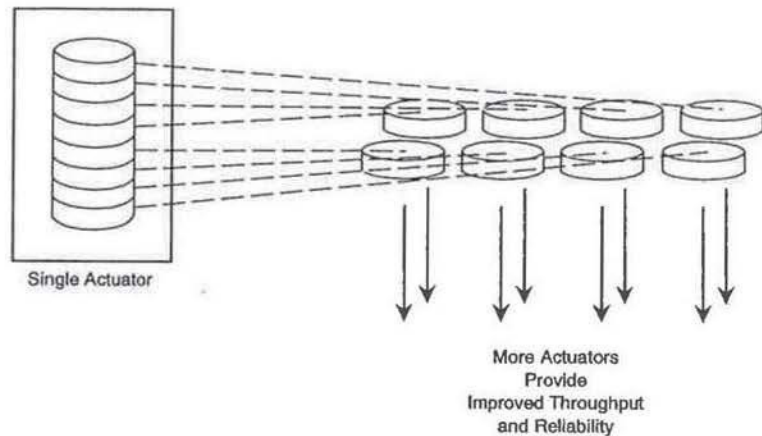


Figure 6.4 Concept of redundant arrays of inexpensive disks.

RAID configurations are as applicable to workstations and LANs as they are to mainframes. While more economical safeguards are available, the mission criticality of information stored on mainframes often makes RAID more relevant for that environment than workstations and LANs.

RAID devices use a software technique called *striping*, which spreads data across multiple disks. Multiple small disk drives operate with a single controller and appear to the host system to be a single logical disk drive.

There are no standards for RAID, but there is general agreement in the industry about the configurations and the functionality associated with each configuration. Generally, the RAID configurations fall into the following categories:

- RAID Level 0 products are technically not RAID products at all, since they do not offer parity or error-correction data, which would provide redundancy in the event of system failure. Although data striping is performed, it is accomplished without fault tolerance. Data is simply striped, block by block, across all the drives in the array. There is no way to reconstruct data if one of the drives fails.
- RAID Level 1 products duplicate the data that is stored on the separate disk drives. This mirroring approach ensures that critical files will be available in the event of individual disk drive failures. Each disk in the array has a corresponding mirror disk and the pairs run in parallel. Blocks of data are sent to both disks at the same time. Highly reliable, Level 1 products are also costly, since every drive requires its own mirror drive, which doubles the hardware cost of the system.
- RAID Level 2 products distribute the code used for error detection and correction across additional disk drives. The controller includes an error-correction algorithm, which enables the array to reconstruct lost data if a single disk fails. As a result, no



expensive mirroring is required. But the code does require that multiple disks be set aside for the error-correction function. Data is sent to one array disk at a time.

- RAID Level 3 products store user data in parallel across multiple disks. The entire array functions as one large logical drive. The parallel operation is ideally suited to supporting imaging applications that require high data transfer rates when reading and writing large files. RAID Level 3 is configured with one parity (error-correction) drive. The controller determines which disk has failed by using additional check information recorded at the end of each sector. However, because the drives do not operate independently, every time an image file must be retrieved all of the drives in the array are used. Other users are put into a queue.
- RAID Level 4 products store and retrieve data using independent writes and reads to several drives. Error correction data is stored on a dedicated parity drive. In level four, data striping is accomplished in sectors, not bytes (or blocks). Sector-striping offers parallel operation; reads can be performed simultaneously on independent drives, thus allowing multiple users to retrieve image files at the same time. However, multiple writes are not possible. The parity drive must be read and written to for each write operation.
- RAID Level 5 products interleave user data and parity data, which are then distributed across several disks. Because data and parity codes are striped across all the drives, there is no need for a dedicated drive. This configuration is suited to applications that require a high number of input/output operations per second, such as transaction processing tasks that involve writing and reading large numbers of small data blocks at random disk locations. Multiple writes to each disk group are possible because write operations do not have to access a single common parity drive.

RAID Levels 1 through 5 enable the array to recover data from a single failed drive, but not from two failed drives. Some vendors are working on additional RAID levels to increase data security and throughput. One of these, RAID Level 6, can recover from a double-disk failure.

For document imaging applications within a work group, a RAID Level 0 configuration might be appropriate, since it supports high data rates for very large files. If continuous availability is a requirement, fault tolerance becomes more important, in which case RAID Level 1 is a better solution. But RAID Level 5 is more economical.

IBM is one of several storage technology vendors supporting the RAID concept. IBM's RAID Level 3 configuration for the RS/6000 workstation, for example, includes as many as 40 5.25-in *enhanced small-device interface* (ESDI) disk drives, which operate at a sustained data rate of 18 MB per second.

#### 6.4 IMPLEMENTATION ISSUES

For many organizations, adding or integrating document imaging applications with existing data processing applications makes good economic sense. However, this decision must rest on other factors as well, such as:



- The degree to which existing applications can benefit from document imaging and the extent to which existing workflows must be redesigned to accommodate a mainframe-based solution;
- The load already being handled by existing data processing resources;
- Availability of spare disk space, memory, and off-line archival facilities to accommodate and protect the image files, as well as the management software;
- Availability of network facilities or spare bandwidth to accommodate the transport of image files between corporate locations and provide alternate routing in case of network failures;
- Availability of in-house staff to develop routines for specialized document processing, using APIs;
- Continuity of data processing staff in following through—from design to implementation of a mainframe-based imaging system.

While adding or integrating image applications to existing data processing applications might seem economical initially, adding the necessary resources to implement document imaging can raise the price considerably. Proper planning will reveal all costs at the start, preventing cost overruns that can damage the credibility of the project and the people involved. The organization might benefit from a workflow study and resource audit, performed by an outside consultant or the professional services group of an imaging system vendor.

## 6.5 CONCLUSION

The mainframe has proven itself a reliable platform for mission-critical applications. Far from being ready to scrap their mainframes, many companies are looking for ways to extend their investments in hardware and software by taking on additional enterprisewide applications. For many companies, document imaging fits the bill.

Incorporating document imaging into the existing host environment can be relatively easy with proper planning. And—compared to the up-front cost of equivalent amounts of processing power, memory, and disk storage in a distributed environment of microcomputers, workstations or both—implementing an imaging system with existing data processing resources can be economical as well.

The mainframe offers the advantage of a single repository where access to sensitive information can be controlled and data can be consolidated, updated, and protected more readily than in a distributed-computing environment. With the aid of CASE tools, 4GLs, and APIs, data processing staff can keep pace with special image processing and management requirements, which, in turn, often translate into competitive advantage. With vendors emphasizing the connectivity and interoperability aspects of their systems, the mainframe is rapidly evolving into an enterprisewide information hub that is fully capable of supporting a wider range of applications, including document imaging.



*Part III*  
**VENDOR IMPLEMENTATIONS**



## Chapter 7

### IBM Imaging Systems

#### 7.1 INTRODUCTION

Despite declining revenues in recent years, a 10-year low stock price, loss of its triple A credit rating and continued staff reductions, IBM Corp. is still the world's dominant computer manufacturer. Its hardware offerings span the three major business platforms: mainframe, midrange and microcomputer. With many business applications moving off the mainframe and onto the other platforms, IBM has had to depend more on the growing outsourcing and systems integration markets, where its revenues grew almost 35% in 1991. IBM also offers a number of software products that together have experienced modest revenue growth. This emphasis on service and software bodes well for IBM's success in the document imaging systems market. Another bright spot in IBM's product portfolio are its token ring networks, which have finally reached market parity with Ethernet.

#### 7.2 IMAGING SYSTEM OFFERINGS

IBM's imaging systems span the three major hardware platforms. The ImagePlus family allows users to tie a number of products together to interact with existing business applications, whether implemented at the work group or enterprise level.

For mainframe environments, IBM offers the *systems applications architecture* (SAA) ImagePlus *multiple virtual storage/enterprise systems architecture* (MVS/ESA), which supports high-volume operational and production applications, providing large organizations with centralized, enterprisewide document, folder, and workflow management.

For midsized businesses or large departments, IBM offers ImagePlus/400 for *Application System/400* (AS/400) environments. The ImagePlus/400 is designed for medium- to high-volume document storage and retrieval, automated workflow processes, and integration with other business applications. The ImagePlus/400 system takes advantage of the IBM SAA ImagePlus Workflow Application Facility/400 and ImagePlus workstation programs, as well as AS/400 hardware and software.

For PS/2 LAN-based environments running *OS/2 Extended Edition (OS/2 EE)*, IBM offers SAA ImagePlus/2. Jointly developed by IBM and Eastman Kodak, ImagePlus/2 is designed to meet low-volume needs of departments in large organizations, as well as the needs of small- to medium-size businesses that handle several hundred documents daily.

Although IBM would prefer that its customers use OS/2 in conjunction with ImagePlus/2, it offers a version of the workstation program that operates under DOS. The DOS version is command driven, whereas the OS/2 version uses Presentation Manager with full SAA-compliant screens. At this writing, IBM has demonstrated a version of ImagePlus/2 that works under Windows 3.1. The primary advantage of using OS/2 is that it offers powerful multitasking capabilities and the ability to window multiple sessions. For many image applications this is an important capability to have.

The components of an ImagePlus system are the same as those found in most other vendors' systems: scanner, display workstations, storage devices, and printer. The system application software implements folder and workflow management. Optional components include facsimile, OCR, and bar code recognition.

For users who want to manage image documents and workflow under IBM's UNIX-based *Advanced Interactive Executive (AIX)* operating system, IBM recommends ImageSystem, which is offered by Image Business Systems, an IBM business partner. ImageSystem works on IBM's RISC System/6000 processors and uses the PS/2 as a high-function image workstation.

A related IBM offering is the ImagePlus *high-volume transaction processing system (HVTPS)*, which provides high-volume image processing for medium to large financial institutions with MVS/ESA environments. Normally, the ImagePlus systems do not read and interpret the contents of documents, but the HVTPS is an exception—it reads written amounts on checks. The product extends image processing to IBM's check-processing control system to provide faster, more efficient financial document handling capabilities.

### 7.3 IMAGEPLUS OVERVIEW

The high-end ImagePlus system operates on IBM System/370 or System/390 processors capable of running MVS/ESA. Four system components run on the host processor:

- Folder Application Facility (optional), which includes the Folder and Workflow Application, Folder and Document Management APIs, and Document Services APIs;
- Object Distribution Manager;
- Data Facility Product (DFP) *Object Access Method (OAM)*;
- Optical storage subsystem.

#### 7.3.1 Folder Application Facility

The Folder Application Facility is a panel-based, menu-driven, front-end application. It automates the traditional systems of filing, retrieving, and processing documents and folders and offers online help.

### 7.3.1.1 Folder Management

The Folder Management function of the Folder and Workflow application maintains on-line information about each document, once the ImagePlus system identifies the document. It also provides an indexing function for images, as well as folder and document security-level assignment functions.

Users can create unique identification schemes for storing scanned documents (image objects<sup>1</sup>) in electronic file folders. They can choose from any number of ways to identify folders, including an account number, customer number, student number, part number, or claim number. In addition, the folders can be categorized by type, such as graduate and undergraduate student folders. The folders can be subdivided with file tabs. For example, a student's folder can be subdivided by such tabs as "Admission and Application," "Medical Records," and "Academic Transcripts," so users can access appropriate records quickly. When the tabs are used to identify and retrieve specific records, they can be displayed side by side with related data files.

With a secondary indexing function, users can consolidate up to three different folder types under one heading, and they can list all folders grouped under a particular secondary index. Users can give each folder type a default value that specifies the folder description and security level. Separate security levels can also be given to each document in a folder.

The Folder and Workflow application provides users with a variety of maintenance options to view and modify image objects and the information about image objects. Folder processing can be suspended at any time, if a user needs to wait until more information can be supplied or additional action taken. When the additional information—a delayed form or signature, for example—becomes available, the user can resume processing at the point at which it was suspended.

The Folder and Workflow application allows batch storage processing of many documents at once. The batch storage process consists of validation, storage, and index updating, as well as document purge or delete from storage. The input to the batch process is usually provided by the high-speed capture subsystem of the Object Distribution Manager, which scans and stores documents in high volumes.

### 7.3.1.2 Workflow Management

The Workflow Management function of the Folder Application Facility automatically routes documents to the various operators, based on categories of documents and definitions. The categories are established in the user profiles created by supervisors at the time the ImagePlus system is configured for operation. When the imaging system is operational, users are only given work that is defined in their profiles. This feature can, for example, prevent inexperienced users from getting high-priority work or prevent users from

<sup>1</sup>An *object* is anything a user can manipulate as a single unit, such as an icon, form, window, text, font, graphic, image, or formatted data.



getting work that is more appropriately handled at a higher level of decision making. The function also tracks the status of work awaiting processing.

The supervisor functions that create user profiles can also be used to grant supervisory authority or folder-routing capabilities to users and to view and manage various work queues. To enhance security, a sign-on feature prevents a single user from signing onto the imaging system at multiple workstations. The system can validate passwords to allow only authorized personnel access to the Folder and Workflow application.

### *7.3.1.3 Application Programming Interfaces*

The optional Folder and Document Management APIs are a collection of programmable commands that provide the functions for indexing, retrieving, and maintaining documents in the ImagePlus system. The APIs support the Folder and Workflow application. They also permit the Folder and Document Management functions to be extended to include the variety of documents that an organization might be storing in other systems or generating in other applications.

The Folder and Document Management APIs provide the following command processing functions to applications developers:

- Folder management commands that create, delete, retrieve, and update folder information;
- Object management commands that manage individual objects and groups of objects;
- Note management commands that add, count, and retrieve folder note entries;
- Object history commands that add, delete, and retrieve object processing events;
- Object pending commands that delete, retrieve, and resend pending object entries;
- Storage management commands that change storage management information.

The Folder and Document Management APIs also have command interfaces that use the Object Distribution Manager to perform synchronous storage and retrieval of objects in the CICS<sup>2</sup> environment.

A migration utility—including source code—is provided with the APIs. This utility automates the conversion of *Data Language 1* (DL/1) databases to DB2 tables, which are used by the Folder Application Facility. DL/1 is the data-manipulation language used in IBM's *information management system/virtual storage* (IMS/VS) environment, whereas DB2 is IBM's relational database management system.

The Document Services APIs provide a set of programming interfaces that allow developers to design applications that are specific to an organization's image processing requirements. The APIs permit the development of a wider variety of specialized, or selective-use, front-end applications.

<sup>2</sup>CICS is IBM's *customer information control system*, a program environment that allows transactions entered at remote computers to be processed concurrently by a mainframe host.

The document services APIs provide the following object command processing requests: store objects, display objects, print objects, and delete objects. They communicate these requests to the Object Distribution Manager.

Distributed with the Document Services APIs (as well as the Object Distribution Manager) is an *Interactive System Productivity Facility* (ISPF)-based application, which provides panels that can be used to maintain the support tables for the Folder and Document Management APIs, the Document Services APIs, and the Object Distribution Manager. This support table administration facility allows the user to view, add, and update data in a support table.

### 7.3.2 Object Distribution Manager

The Object Distribution Manager works as a file server for image objects and for nonimage objects, such as coded data. Positioned at the functional center of the ImagePlus system, the Object Distribution Manager coordinates the activities of the system's major components. It receives new image objects from the scanner-equipped user workstations and routes these objects to the OAM or to a sequential file for future batch or coded-data storage processing. The Object Distribution Manager also receives user requests for display, print, and other object processing functions. It passes these requests to the system's storage components, receives the requested objects, and routes the objects to the appropriate workstation or LU 6.2 application.

In conjunction with the APIs, the Object Distribution Manager supports the generation of the object name, collection names, user transaction handling reference, and symbolic host/symbolic workstation reference.

### 7.3.3 Object Access Method

The OAM provides the means for moving data between the peripheral devices and the host operating environment. By supplying storage criteria for the types of documents that will be processed and the rules that must be used in processing them, the OAM provides system-managed storage for objects and allows the storage management cycle to be controlled by the user. The OAM then automatically stores, retains, and subsequently migrates and deletes objects according to user-defined document processing cycles.

The OAM uses the storage criteria to store objects in a storage hierarchy consisting of a high-speed magnetic disk and a high-capacity optical disk. It maintains an object directory, which allows access to every object stored in the system, regardless of where in the storage hierarchy the object currently resides.

The OAM stores new objects and frequently accessed objects on high-performance magnetic *direct-access storage devices* (DASD), such as the IBM 3390. It migrates objects to an optical disk library when the objects reach the user-defined migration criteria. Image document retrieval from DASD takes from four to six seconds, whereas an image document retrieved from the optical disk library takes approximately 30 seconds.

If the document processing cycle includes the eventual deletion of archived material, the OAM simply erases the expired object's address information, resulting in its logical deletion. The OAM also supports the option of concurrently writing the object to a backup optical disk, which IBM calls *image shelf storage*.

The OAM tracks all magnetic and optical storage disks on the ImagePlus system, including all library and shelf optical disks. It also directs the mounting and removal of these disks, in answer to user needs, and the staging of objects between magnetic and optical storage, for enhanced response time.

The IBM 9246 Optical Library Unit is the centerpiece of the optical storage system. It includes FileNet's Optical Disk Storage and Retrieval Library Unit (OSAR 64 Model 1100) and two *Laser Magnetic Storage International (LMSI) LaserDrive 1250E* optical disk drives. The library unit also contains a mechanical accessor, to retrieve and mount the optical disks as needed, and storage slots for up to 64 disks. For large organizations, the library unit can be expanded to contain four disk drives.

For storage media, the IBM 9246 uses LMSI LM1200 optical disks. These are 12-in WORM optical disks that are capable of storing 128 GB of information on line, which is equivalent to over 2,500,000 page images at a typical size of 50 KB per page.

Figure 7.1 shows a typical IBM SAA ImagePlus MVS/ESA system and illustrates the various component relationships.

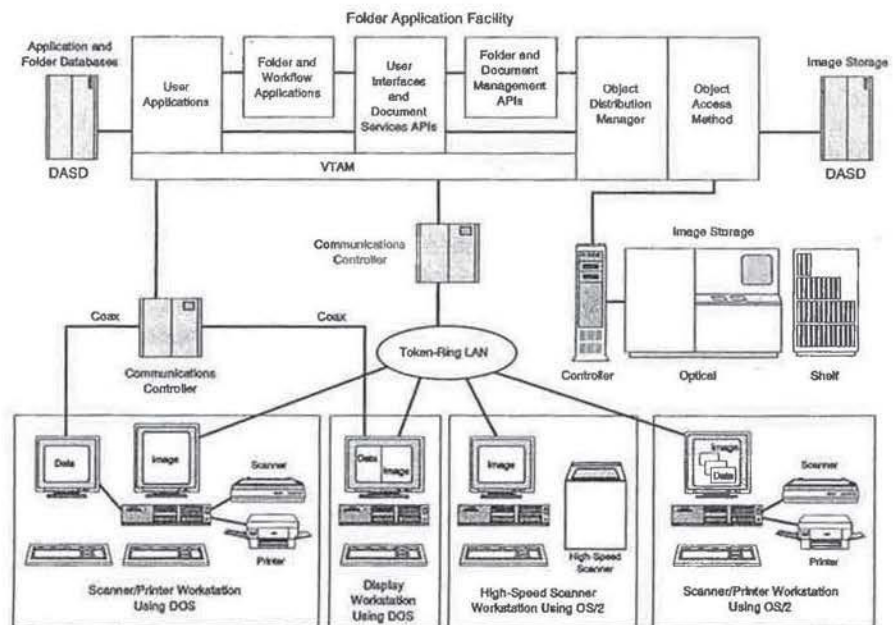


Figure 7.1 IBM SAA ImagePlus MVS/ESA system. Source: IBM Corp.



When connected to a token-ring LAN, an image workstation requires two types of cards: an image adapter card and a token-ring card. The image adapter card performs the image compression/decompression processing, to conserve network bandwidth. However, a recent software upgrade now allows compression/decompression to be handled by the workstation's CPU, eliminating the need for a separate card—provided that the workstation is a PS/2 Model 70 or higher. The token-ring card is used to establish an LU 6.2 session over the LAN between the workstation and the host. If the workstation is used to support 3270 sessions, a terminal emulation card is also required.

#### 7.4 DOCUMENT ENTRY

As with any other imaging product, entering paper documents into the ImagePlus system involves two steps: indexing and scanning. These steps can be performed by using a manual process for each page or by using a high-speed scanner to enter large volumes of documents. With high-speed scanning, the batch storage process can be used to validate document indexing information, store the documents, and update the Folder Application Facility index and routing information.

The ImagePlus offers a number of scanners, including flatbeds and automatic sheet feeders from such vendors as Ricoh, Howtek, Bell and Howell, and Kodak. IBM's high-speed capture subsystem uses the Kodak ImageLink Scanner 900. This unit is equipped with a paper stacker for automatic feed and performs single-pass transport and duplex scanning, bar code reading, and page imprinting. The Folder Application Facility batch storage process uses output from the scanner to index and store documents and optionally route them for processing through workflow queues.

With the manual process, index information is entered using Folder and Workflow application panels. The system provides a temporary ID that uniquely identifies the document and associates the document with the folder into which it is to be placed. Then the document is scanned into the system at a scanner-equipped user workstation. The manual process is shown in Figure 7.2.

When documents arrive in the mail, they are taken to an operator workstation where the Folder and Workflow application resides. The operator uses this terminal to index the documents. The operator enters descriptive information about the documents and relates them to one or more electronic file folders.

At this point, the operator can also specify how the new documents are to be routed. The routing information might be indicated by the form number entered for the document. For example, a form number might indicate what department, work group, or individual should get the document first. The operator submits the indexing information to the Folder and Workflow application, which returns a temporary ID. This number is kept with the hard copy of the document. The document is then indexed and associated with a folder, and can be taken to a scanner-equipped workstation.

After the temporary ID is keyed in at the scanner-equipped workstation, each page of the document is scanned into the ImagePlus system. The scanner operator checks the image display to ensure that the resulting image is clear and legible. The collection of

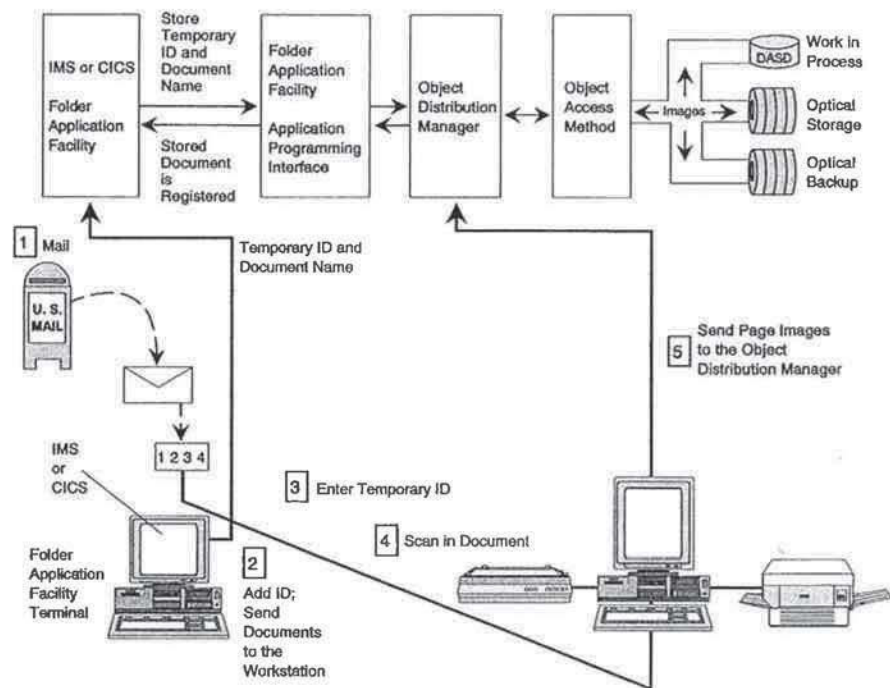


Figure 7.2 Manual indexing and scanning processes. Source: IBM Corp.

scanned-page images is then submitted to the IBM SAA ImagePlus Object Distribution Manager MVS/ESA as an image object.

Within the ImagePlus system, the Object Distribution Manager receives the new image object and routes it to the OAM, which stores the new object on magnetic disk. The Object Distribution Manager also notifies the Folder and Workflow application that the document with the temporary ID has been scanned and stored in the OAM.

With the image object now registered in the system, the Folder and Workflow application places the object in a routing queue for subsequent routing to an appropriate user via the Workflow Management function. The Folder and Workflow application then updates the indexes and makes the object available for processing.

## 7.5 WORKFLOW

Users can process documents that are assigned specifically to them, or they can process work that is defined in their user profiles. In a standard request the user simply issues a command instructing the system to get work and a document is provided for processing.

The Folder and Workflow application determines the kinds of objects the user is authorized to process and selects from this group the object with the highest priority. In a request for a specific object the user views the documents of a folder and selects the desired item. For both types of requests the object and any related data are retrieved and sent to the workstation. This process is shown in Figure 7.3.

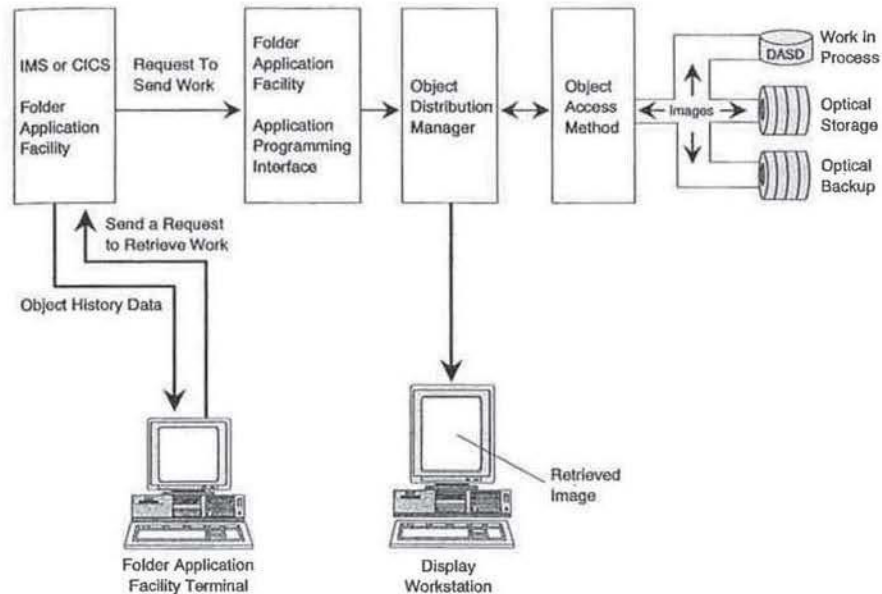


Figure 7.3 Retrieving a document for processing. Source: IBM Corp.

The first page of the selected object appears on the workstation's image display monitor. The user is able to scroll through and manipulate the displayed image using the functions provided by the workstation program. The object's processing history can be displayed on the workstation's data display monitor at the user's request. This processing history presents an audit trail of all work performed on the object up to that point. The supervisor also has access to this information, independent of any workstation operator.

Within the Folder and Workflow application objects are either in the routing queue or filed. Objects are filed if they do not need processing or if processing is complete. Filed objects can be put back into the routing queue upon command by a workstation operator or can enter automatically when it is time for processing.

Generally, users will process and then file documents, but they also have the option of holding a document, which delays further processing until a specified date and time. Documents can also be rerouted to other users on the ImagePlus system. Sometimes the supervisor will reroute documents to balance the workload among operators.



The Folder and Workflow application helps users track their assignments. Users can call up a list of documents currently assigned to them or all of the documents belonging to a specific folder currently being routed. When viewing this information users can access the processing history of a document and perform maintenance on the document. Users can manage processing of the document by reassigning it to the person who handled it last or by suspending it for later processing, using the hold capability.

Based on the information entered during indexing, documents can be prioritized for processing. A base priority value is set and incrementally raised each day the document remains unprocessed. When work is requested by a workstation user, the system responds with the highest priority document that user is qualified to work on. This helps assure that work is distributed evenly and that the most critical work is handled first.

The workflow management function can also handle the unpredictable processing requirements of a business. For example, when it is discovered that a document being processed is missing key information, it can be turned back to the previous workstation to have additional work performed before it is processed any further. To accomplish this, the document can be changed from one routing queue to another or assigned specifically to a user for priority processing.

Workflow management provides the best illustration of productivity improvement in the office environment. In large organizations there are tens of thousands of paper documents processed daily, most of them scattered across desks in hundreds of different offices and cubicles. It is impossible to measure the work waiting to be done, let alone check on its progress. With the workflow management software of an enterprisewide imaging system, each document can be individually tracked, assigned a priority for control, and selected for processing from input queues. This allows supervisors to measure the productivity of workers with statistics generated by the system, which in turn allows management to more accurately forecast staffing requirements and make other workload-related decisions.

## 7.6 FOLDER MANAGEMENT

Documents can be indexed into folders after scanning. Up to three secondary indexes can be assigned to a folder or group of folders to provide alternative means of identification. For example, an employer insurance policy number can be used to group all employee claims for that policy. The workstation user can display a list of folder contents and select a particular document for viewing or printing. Likewise, a list of folders grouped under a particular secondary index can be listed, a folder selected, and a list of its contents displayed. Documents can be moved from one folder to another.

Folders can be created with attributes such as folder type and security level. These attributes can be updated as needs change. Folders can also be created with temporary IDs that can be changed later to permanent IDs. This is useful for a bank or mortgage company, when, for example, customers apply for loans. The folder can be assigned a temporary ID and, when the loan is approved, a permanent ID can be assigned.