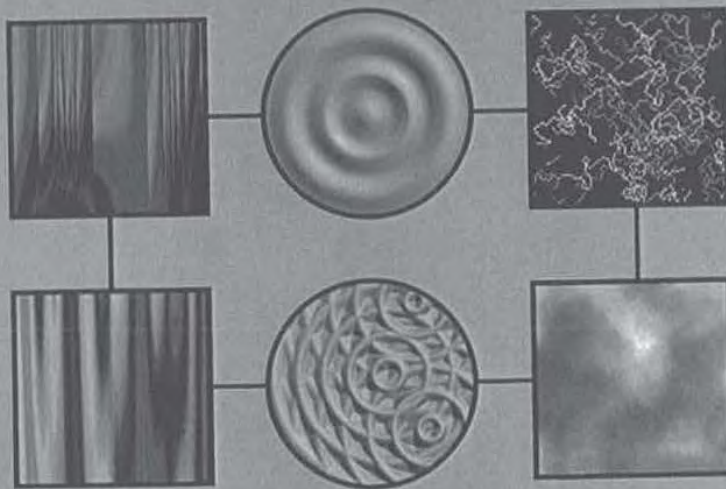


# Networks and Imaging Systems in a Windowed Environment

MARC R. D'ALLEYRAND  
EDITOR



# **Networks and Imaging Systems in a Windowed Environment**

*The Artech House Computer Science Library*

*ATM Switching Systems*, Thomas M. Chen and Stephen S. Liu  
ISBN: 0-89006-682-5

*Client/Server Computing: Architecture, Applications, and Distributed  
Systems Management*, Bruce Elbert and Bobby Martyna ISBN:  
0-89006-691-4

*Computer-Mediated Communications: Multimedia Applications*,  
Rob Walters ISBN: 0-89006-757-0

*Computer Telephone Integration*, Rob Walters ISBN: 0-89006-660-4

*Distributed and Multi-Database Systems*, Angelo R. Bobak  
ISBN: 0-89006-614-0

*A Guide to Programming Languages: Overview and Comparison*,  
Ruknet Cezzar ISBN: 0-89006-812-7

*Introduction to Document Image Processing Techniques*,  
Ronald G. Matteson ISBN: 0-89006-492-X

*Managing Computer Networks: A Case-Based Reasoning Approach*,  
Lundy Lewis ISBN: 0-89006-799-6

*Survival in the Software Jungle*, Mark Norris ISBN: 0-89006-831-3

*UNIX Internetworking, Second Edition*, Uday O. Pabrai  
ISBN: 0-89006-778-3

*Wireless LAN Systems*, A. Santamaría and F. J. López-Hernández  
ISBN: 0-89006-609-4

*Wireless: The Revolution in Personal Telecommunications*, Ira  
Brodsky ISBN: 0-89006-717-1

*X Window System User's Guide*, Uday O. Pabrai  
ISBN: 0-89006-740-6



# **Networks and Imaging Systems in a Windowed Environment**

---

---

Marc R. D'Alleyrand, Ph.D.  
Editor

Artech House  
Boston • London



**Library of Congress Cataloging-in-Publication Data**

Networks and imaging systems in a Windowed environment/Marc R. D'Alleyrand,  
editor

p. cm.

Includes bibliographical references and index.

ISBN 0-89006-654-X (alk. paper)

1. Image processing. 2. Computer Networks. 3. Windows (Computer programs)

I. D'Alleyrand, Marc R.

TA1637.N45 1995

005.75-dc20

95-42036

CIP

A catalogue record for this book is available from the British Library

© 1996 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.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 0-89006-654-X

Library of Congress Catalog Card Number: 95-42036

10 9 8 7 6 5 4 3 2 1

# Contents

Preface xv

## Part I Introduction to Networks and Imaging Systems 1

---

### 1 Networks and Imaging Systems in the Business Process 3

- 1.1 Introduction 3
- 1.2 Orders of Magnitude and Other Considerations 4
- 1.3 Evolution of Networks 5
- 1.4 Evolution of Processors 7
- 1.5 Evolution of Operating Systems and Environments 8
- 1.6 Evolution of User Interfaces 9
- 1.7 Implications for the Evolution of Imaging Systems 10
- 1.8 Implications for Workflow Reengineering 12
- 1.9 Conclusion 13

### 2 Introduction to Networked Imaging Systems 15

- 2.1 Introduction 15
- 2.2 Imaging Systems as Information Transfer Agents 18
- 2.3 Industry Trends 18
- 2.4 The Role of Imaging in Information Transfer 20
- 2.5 Various Types of Imaging Systems 21
  - Front-End Applications 21 • Management Applications 21 • Archiving Applications 22
- 2.6 Technological Concerns 23
- 2.7 Implications for Systems Architecture 24
- 2.8 A Multilevel Network Functional Architecture 25
  - Functional Factors Influencing Imaging Network Design 25 • The Different Levels 27



vi *Contents*

2.9	Implications for User Interfaces	29
	Image Display	29 • Commands Access 29 • Controls and Management 29
2.10	Structure of an Optical Disk System	30
	Document Capture	30 • Processing 30 • Storage 31 • Display 31 • Communication 31
2.11	Examples of Applications	32
2.12	Design Considerations	33
2.13	Implementing an Optical Disk System	35
2.14	Conclusion	36
<b>3</b>	<b>Imaging Networks</b>	<b>37</b>
3.1	Introduction	37
3.2	Network Fundamentals	39
	Network Architecture	39 • Defining a Network 40 • Network Components 50
3.3	Selecting an Imaging Network	59
	Conducting a Needs Assessment Survey	60 • Selecting a Workstation 62 • Selecting a Document Grouping Structure 64 • Selecting a Network Topology 65 • Selecting Operating Systems and Protocols 65 • Selecting Servers 68
3.4	Other Requirements	70
	Network Management Software	70 • Network Simulation 70 • Configuration Management and Version Control 70 • Workflow Management 71
3.4	Conclusion	71
	Notes	71
<b>4</b>	<b>User Perspective of Imaging Systems</b>	<b>73</b>
4.1	Introduction: The Business Case	73
4.2	The Components	74
4.3	The Workstation	76
	The Desktop	77 • System Unit 78 • The Rest of the Hardware 83 • Software 88
4.3	Conclusion	90



**5 Considerations for Network Imaging Hardware 91**

- 5.1 Introduction 91
- 5.2 Scanners 91
  - Transport Mechanism—Paper Scanners 92 • Scanners—Microforms 94 • Characteristics of Scanners 94 • Scanner Adjuncts 94 • Fax Machines as Scanning Devices 95
- 5.3 Printers 95
  - Printers and Print Servers 95 • Different Types of Imaging Printers 96 • Network Printing Issues 98 • Raster, PostScript, TrueType, and Other Printing 98
- 5.4 Document Storage 98
  - Hard Disks 99 • Optical Disks 99 • Tape Storage 99
- 5.5 Conclusion 99

**6 Display Selection for Document Imaging in Windowed Environments 101**

- 6.1 Introduction: Display Requirements in Windowed Environments 102
- 6.2 Image Readability 103
  - Research on Readability and Eye Fatigue 103 • Display Factors Affecting Readability and Eye Fatigue 105 • VGA Problems 111
- 6.3 Display Performance 111
  - Document Image Scaling and Decompression Alternatives 111 • Windows and Document Image Processing Display Performance 114 • The Cost of Slow Image Flip Rate 114
- 6.4 Selecting the Right Display 115
  - Displays for Dedicated Document Imaging Systems 116 • Document Imaging in a Mixed Application Environment 116 • Impact of Windows on Display Selection 116
- 6.5 Conclusion 118

**Part II Planning and Design 119****7 Planning an Imaging System: Workflow Redesign 121**

- 7.1 Introduction 121

7.2	Guidelines for Selection of a Networked Imaging System	122
	System Architecture	123 • Data Organization 123 • Platform Selection 124 • Indexing and Data Identification 124 • Integration of the Strategic Decisions 125
7.3	Basic Principles and Objectives of Workflow Redesign	125
	Introduction	125 • Objectives of Workflow Redesign 127 • The Redesign Process 127
7.4	Workflow Redesign	128
	Present System Analysis	129 • Identification of Required Changes 130 • Tasks 130 • Performance 131
7.5	System Redesign	131
	Addition	131 • Deletion 132 • Consolidation 132 • Redistribution 132
7.6	Selection of a Technical Solution	133
7.7	Implementation	133
	Design of the Implementation Plan	134 • Determination of Acceptance Criteria 134 • Procurement 134 • Training 134 • Conversion 134
7.8	Verifying the Redesign	135
	Workflow Simulation	135 • Network Simulation 137
7.9	Strategic Decisions for Workflow Redesign	137
	Adoption of Standards	138 • Data Capture 138 • Indexing 139 • Information Retrieval 141 • Processing System 142 • The Data Networking Environment 142
7.10	Conclusion: Redesign and the Organizational Structure	142
	Notes	143
<b>8</b>	<b>Justifying a Networked Imaging System</b>	<b>145</b>
8.1	Introduction: The Justification Process	145
8.2	Limitations of Traditional Justification Models	146
8.3	Defining a Justification Model for Imaging Applications	146
	Determining the Cost Structure	147 • Building a Financial Justification Model 147 • Cost Components of a Financial Justification 148
8.4	Life-Cycle Justification	156
	Project Justification and Project Survival	157
8.5	Examples of Cost-Justification Activities	157
	Chronology of the Justification Process	159 • Determination of Costs and Benefits 161 • Preparation of the Justification Document 163



8.6	Methodology for Selection of Network Document Imaging Applications in a Competitive Environment	164
	Ranking Method	166
8.7	Conclusion	167
<b>9</b>	<b>Image-LAN Enabling and Integrating LAN-Based Imaging Systems with Existing Mainframe Applications</b>	<b>169</b>
9.1	Introduction	169
9.2	Current Trends and Challenges	171
9.3	Impact on Traditional Transaction Processing Operations	172
9.4	Moving from Low- to High-Volume Transaction Processing	173
	Low-Volume Transaction Processing	173 • High-Volume Transaction Processing
	Matching Imaging System Features with Mainframe Requirements	174 • Examples of Transaction Types and Processing
	Measuring Results	175 • Document Image Caching, Routing, and Staging in a High-Volume Environment
9.5	Productivity Factors	178
	Human Factors	178 • Technical Factors
9.6	Issues in Implementing Image Enabling	179
	Workflow Control	179 • Checklist of Technical Issues
	Standardized Integration Tools: Imaging and Mainframe Applications Working Together	181 • Front-End Tools and Workstation Issues
	LAN/WAN Integration for Dispersed Geographic Sites	183 • Backup and Recovery
	Additional Consideration When Downsizing	185 •
9.7	Conclusion: Outline of Implementation Strategy	186
<b>10</b>	<b>Imaging Data Structure and Indexing</b>	<b>189</b>
10.1	Computer Images	190
	Binary	190 • Gray Scale
	Color	191 • Video
10.2	Compressed Image Formats	191
	Plotter/Vector Descriptors	192 • Compressed Binary
	Compressed Gray Scale and Color	193 • Compressed Video
10.3	Image Storage and Retrieval	194
	Electronic Document Filing	197 • Electronic Image Filing



- 10.4 Image/Document Database Index Design 199
- 10.5 Indexing Moving Images 202
- 10.6 Advanced Image Databases 202
- 10.7 Index Data Structures 203
  - Image Data 203 • Structured Control Information 204 •
  - Structured Control Information Access Indexes 204 •
  - Unstructured Text Descriptions Data 204 • Unstructured Text
  - Description Content Indexes 205 • Format Information 206
- 10.8 Geometric Description Objects 206
- 10.9 Object Recognition Systems 208
- 11 Database Structure for Document Image Retrieval 213**
  - 11.1 Elements of the Imaging (Object) Database 215
    - Introduction 215 • Database Models for Imaging
    - Applications 215 • Database Structure for Document
    - Retrieval 218 • Archival, On-Line Reference, and
    - Transactional Applications 219 • Indexing Methodology 219
  - 11.2 Design Requirements of the Imaging Database 222
    - Overall Considerations 222 • Other Considerations: Cultural
    - Elements of Systems Design 223
  - 11.3 Preparing the Imaging Database—Preliminary Questions 225
    - What Information Is to Be Stored as Images? 225 • How Is
    - the Information Currently Stored? 226 • How Much
    - Information Is to Be Stored? 228 • What Kind of Retrieval
    - Is Needed? 229
  - 11.4 Preparing the Imaging Database—Operational Questions 230
    - How Will the Images Be Captured? 230 • How Will Data
    - Be Extracted from the Images? 232 • What Structure Will
    - Be Applied to the Data? 233 • Where Will the Image Be
    - Stored? 234 • Media Selection 235
  - 11.5 Quality Control 235
  - 11.6 Conclusion: Image Enabling Existing Installations 236
  - Notes 238
- 12 Graphical User Interfaces and Imaging Systems 239**
  - 12.1 Structure of Standard GUIs 240
  - 12.2 An Overview of GUI Services 242
    - Metaphors to Compute By 242 • Windowing Models 244 •
    - Overview of Operating and Windowing Systems 246

- 12.3 Issues in GUI Development 248
  - Portability for a Network Environment 249 • Performance for Image Rendering 251 • Performance for Event Handling 252 • Performance for Cooperative Processing 253 • Localization 255
- 12.4 Multimedia and Metaphor Shifts 256
- 12.5 Conclusion 257

## Part III Implementation 259

---

### 13 Image Management Development Software 261

- 13.1 Introduction 261
- 13.2 Selection Criteria 262
- 13.3 Strategies for Implementing Image Management Software 264
- 13.4 Impact of Network Computing Developments 265
- 13.5 Comparing Commercial Products 266
  - Existing Platform Support 266 • Image Indexing and Classification Methods 266 • Database Independence 266 • Document Formats 267 • LAN Support 267 • Host Application Integration 267 • Support for Existing Applications Programs 268 • Image I/O Peripheral and Optical Storage Device Support 268 • Application Interfaces to Imaging Services 268 • System Administration Facilities 268
- 13.7 Conclusion 269

### 14 Tools and Toolkits for Network Imaging and Workflow Systems 271

- 14.1 Introduction 271
- 14.2 The NIWS Difference 272
  - The Lay of the LAN—Open Client/Server Applications 272 • Compound Data over the Network 272 • Sophisticated Work Processes 273
- 14.3 Technologies That Make a Difference 273
  - Imaging as a Basic Technology for Work Process Automation (WPA) 273 • Optical Storage 273 • Graphical Environments 274 • Imaging and the Database 274 • Optical Character Recognition 274



xii *Contents*

- 14.4 Work Process Automation and Business Reengineering 275
  - Ad Hoc versus Structured 275 • Interactive Workflow Management 276 • Using Workflow to Integrate Compound Data 277
- 14.5 WPA Tools 277
  - Choosing Your Weapons: The Pro-Am 277 • Criteria for Toolkits 278 • Workflow Management Tools 280 • Scripting versus Graphical Tools 280 • Examples of Workflow Tool Capabilities 281
- 14.6 Factors in Designing Workflow Systems 282
  - Processing Activities 282 • Input Devices 283 • Input/Output to Workflow Activities 283 • Data Access Considerations 283 • Multiple Server Considerations 283 • Processing Time for Each Activity 284 • Processing Hardware Requirements 284 • Workflow Mapping Features 284
- 14.7 Workflow Automation and Imaging Tools 286
  - Flexibility and Ease-of-Use 286 • Transparency 286 • Openness and Adherence to Standards 288 • System Architecture and Design 288 • Application Database Management Considerations 290
- 14.8 Implications for Optical Storage 291
  - Caching 291 • Prefetching 291 • Staging 292
- 14.9 Enabling Extant (Legacy) Systems 292
  - Dilemma: Rest or Recreation 292 • Host-Based Applications: A Terminal Condition 292 • Data Exchange 293 • Workflow Gateways with Heterogeneous Systems 294 • Using CASE with Imaging Application Tools 295
- 14.10 Workflow System Application Development Methodology 295
  - Structured Methodology as Language 295
- 14.11 Conclusion 300
- 15 Integrated Workgroup Image-Enabled Productivity Software 303**
  - 15.1 Definition 304
  - 15.2 The Image as the Application 305
  - 15.3 The Image as a Component of the Application 306



- 15.4 Design Considerations for an IWIPS 307
  - Define Business Operation 307 • Windows Workgroup Imaging System Architecture 308 • Client/Server Computing 309 • Features for Integrated Workgroup Software 313 • Printing Document Images 315 • Selection Criteria for Integrated Workgroup Software 316
- 15.5 Some Final Points to Consider 316

## 16 Combining Microforms and Optical Disks 319

- 16.1 Introduction 319
- 16.2 Generating MDIs 321
  - Traditional Method: Film-Then-Scan 321 • Concurrent Method: Scan-and-Film 323 • Advanced Method: Scan-Then-Film 325
- 16.3 Comparison between the Three Methods 326
  - Quality 327 • Image Sharpness and Resolution 327 • Operational Considerations 329
- 16.4 Cost Comparison between the Various Methods 331
  - Equipment 331 • Personnel Costs 331
- 16.5 Other Considerations 332
  - Equipment Sharing 332 • Database Reconciliation 333 • Records and Archive Management 334 • Data Extraction 335 • Document Composition 335 • Selection of Best Method 335
- 16.6 Application to Cold Technologies 335
- 16.7 Conclusion 336

## 17 Deploying an Imaging Network 337

- 17.1 General Systems Deployment Considerations 338
  - Data Storage Considerations 338 • Retrieval Times Impact 339 • Optimized Use of Resources 339 • Centralized versus Distributed Criteria 339 • Legacy Systems Integration 340
- 17.2 Integration with Current Systems 340
  - PCs, Minis, Mainframes 340 • SNA Wide Area Networks 341 • Token Ring and Ethernet 341

xiv *Contents*

- 17.3 Network Infrastructure 341
  - Low-Speed Network Overhaul 341 • Meshed Architecture 342 • Advanced Bridges and Routers 343 • Network Management 344 • High-Speed Transports 345 • Traffic Modeling 345
- 17.4 Personnel Ergonomics 346
  - Workstation Platforms 347
- 17.5 System Structure 348
  - Distributed Theory 349 • Centralized Theory 350
- 17.6 Conclusion: The Choke Points 352

**Index 355**



# Preface

Over the past few years, imaging has grown from a specialized, stand-alone technology, used as an alternative to microforms, to a powerful tool in the design of information systems, bridging the operational and intellectual differences between machine-readable and non-machine-readable information. However, this change in emphasis has not changed the technical requirements that govern the implementation of an imaging system. In particular, imaging systems still require the articulation of two separate processes: one covers information capture, processing, storage, retrieval, and display of documents; the other governs the electronic connections between the imaging system and its users.

Integrating an imaging system in an existing organization must often be done in the context of an existing computer system and of a network designed to handle the data necessary for the support of an organization's information needs. Two approaches are usually contemplated:

- Image enabling, which attaches images of documents to computer applications
- Image integration, which allows interaction between the computer application and the document contents

In most cases, image enabling does not result in significant changes in electronic data flow. This is not the case with image integration. However, in either case, because efficient imaging systems require a significant amount of resources, the addition of imaging capabilities to existing systems should not be considered as a mere addition of another application. Handling and processing images requires much larger computing capabilities than are needed for traditional data. Consequently, adding imaging to an existing installation will demand a complete reassessment, and often redesign, of the existing computer and communications systems. In all but the simplest cases, image enabling will lead to a reevaluation of the best workflow to support an organization's business objectives, often resulting in a complete redesign of the business process itself.

This reevaluation will identify systems requirements associated with the imaging capabilities and provide guidance for selection and design of image-based



information systems. It will also analyze the impact of imaging on existing systems and provide directions for reengineering a conventional information transfer process to take advantage of an imaging environment.

It is the purpose of this book to present the discussion of design and implementation concepts made possible by the availability of multitasking (windowed) and communicating (networked) technologies and to identify issues and solutions. As such, it is a follow-up to two earlier books that dealt with the why and how of imaging.<sup>1</sup> This work focuses on the workstation as the physical interface between the organization and its information system and on the constraints that the use of imaged documents places on such an interface.

Additionally, this book is designed to be a guide to the selection, planning, design, and integration of image-based document storage and retrieval applications. The emphasis will be on printable/displayable text and graphics materials. However, its approach will be applicable to the emerging field of multimedia information systems, using video and voice in electronic document systems.

In particular, this book will also focus on the impact that the introduction of new processing chips, such as the Pentium®, and of new operating systems, such as Windows95® and Windows NT®, may have on imaging and document systems.

The scope of this book is:

- The identification of the critical issues that are necessary to the efficient use of the technology in a multitasking environment.
- The description of the tools supporting that environment.

*Networks and Imaging Systems in a Windowed Environment* is written for the end users of document retrieval systems. The extension of electronic imaging beyond traditional document processing tasks makes it valuable as a vehicle for interdisciplinary development. Written with minimum technical terminology and from the point of view of the end user, this book will be of interest to designers of office systems, who need to understand the constraints of imaging systems; end users, who need a frame of reference to translate their needs into specifications; system planners and implementers, as well as support personnel, who need to understand the practical aspects of installed systems; and students in the field, who need comprehensive treatment of specific applications of the technology. Because of its contents, this book also will likely be of interest to technical personnel involved one way or the other in the design and planning of imaging systems, as it will clarify some of the issues facing their clients.

Please note that throughout this book, the term *data* is often used to refer to machine-readable signals, and the term *image* is used to describe the digitized representation of a paper or microform document.

This book has been assembled with great care. However, because of the complexity of the field and the fact that there is no single or even best technical solution, there is no guarantee that a particular recommendation will work in a

specific environment. Consequently, the contributors to this work, as well as this editor, cannot be held responsible or liable for any negative outcome of their writings. In addition, while occasional repetition of subject matter between chapters has been avoided as much as possible, it has sometimes been necessary to reiterate earlier statements.

—Marc R. D'Alleyrand

---

## NOTES

---

- <sup>1</sup> D'Alleyrand, M. R., *Image Storage and Retrieval Systems* (New York: McGraw-Hill), 1989; M. R. D'Alleyrand, ed., *Handbook of Image Storage and Retrieval Systems* (New York: VanNostrand & Reinhold), 1992.



---

# 2

---

## Introduction to Networked Imaging Systems

---

*Marc R. D'Alleyrand, Ph.D.*

---

### 2.1 INTRODUCTION

---

Within a few years, imaging systems have outgrown their role as “electronic filing cabinets” to become necessary components in any business system that involves the handling of documents in a nonelectronic medium. This, in turn, has required integration of imaging systems with the existing organizational computer infrastructure, as well as the integration of the imaging functions with the business applications.

These integrated systems must resolve the same issues as their “imaging only” predecessors, for example:

- *Operational*—to eliminate the operating difficulties associated with the unnecessary handling of vast amounts of paper.
- *Financial*—to reduce the costs associated with the finding and retrieval of information in an expanding database and filing system, as well as costs resulting from printing and copying.
- *Security*—to increase ease of access to authorized information, without jeopardizing security and confidentiality requirements. Of all the issues raised, this one is potentially the most troublesome to resolve, because the increased sophistication of network penetration techniques may compromise a secure system overnight.

- *Dissemination*—to ensure that the information is captured, processed, stored, displayed, and communicated under optimal conditions to the right people and in a timely manner.

This must be done in an environment requiring faster access to information contained in the documents.

Earlier document systems were inefficient due to the need to maintain separate filing systems: one for non-machine-readable documents (e.g., paper folders, microfilm rolls, microfiches) and one for the information about these documents or the information they contained (e.g., machine-readable computer reports and indexes). Adding to this inefficiency was the inability to transmit non-machine-readable documents electronically and/or the need to create multiple copies to satisfy concurrent access requirements.

The development of optical disk technology, with its large storage capacity, has made it practical to store both documents and information about them on the same medium in a digital form. This can be done regardless of content, be it machine-readable text, graphics, sound recordings, or images; and regardless of original format, be it paper, microform, tape, or other medium. To give an example: A high-density 5.25-inch diskette, with a storage capacity of 1.2 million bytes, can store approximately 500 one-page typewritten memoranda in a machine-readable format. That same diskette cannot store the uncompressed image of one page. Even with most common compression schemes, such a disk would probably be able to store only ten to fifteen images of typewritten text—a number that rapidly decreases if the document contains shades of gray, such as in a photograph. By comparison, a 5.25-inch optical disk platter is capable of storing 20,000 pages, and a 12-inch platter can store up to 50,000 pages in a digitized image format.

Furthermore, because the storage process involves computer encoding, the information can be manipulated as computer data. In addition, digital signal processing is capable of electronic image enhancement, which means that it is possible to recover documents that, under conventional optical methods, would be considered unusable. This is of specific importance when attempting to capture images of marginal material, such as poor carbon copies, poorly recorded microforms, or decaying paper with faded ink.

Optical disks make it possible to access, disseminate, and interpret document images in a manner similar to simple machine-readable material, especially in terms of security and encryption. That technology is of particular value to office correspondence, where documents such as personnel files and other administrative materials are created on paper in a non-machine-legible format. Because it stores not only text but images, the digital format makes it possible to retrieve, disseminate, and display items such as facsimile signatures and handwritten marginalia in ways that were not possible before.

The combination of imaging and data processing in optical disk-based systems results in the following characteristics:



- *Immediacy* Electronic capture of images eliminates the need for further manual processing of the original documents. Consequently, access to electronic images and updates is practically instantaneous, giving the ability to maintain files in a timely fashion.
- *Communication* Because the electronic image is kept in a digitized form, it can be transmitted through data communication lines using well-known computer techniques. Encoding and standards might differ between systems, however, and will have to take into consideration the large amount of data being communicated. Nevertheless, the increased acceptance of protocols, such as those recommended by the International Telephone and Telegraph Consultative Committee (CCITT), facilitates the design of compatible networks.
- *Image enhancement* Image processing algorithms make it possible to detect edges of images and modify contrasts and thresholds, thereby reclaiming images that would have been unreadable in a microfilm environment. It is possible, for example, to strengthen the contrast in a blue carbon copy on a pink paper, an impossible task with microfilm.
- *Data extraction* The development of affordable image processing software has made it possible to incorporate automatic data extraction in the imaging workflow and to integrate imaging into the business process. Because of the similarity of signals generated by an image scanner and an optical character reader (OCR), it is possible to process the image through an appropriate program, which will identify and extract machine-readable characters in that image, thus enabling the capture of an image and extraction of the data that it contains at the same time.
- *Multimedia recording* Electronic images are only one of the information elements that can be captured on optical disks. Other information elements, such as voice and music, can, once digitized, be added to the electronic records, making it possible to have an image stored with a verbal comment.
- *Hypermedia* One capability offered by electronic imaging is called hypermedia, which is an arrangement of related documents where cross-referenced information can be accessed from any of the documents without having to leave the first document. Early examples of hypermedia, called hypertext, are well known, the simplest being a footnote or an explanation in the margin. What is important for information retrieval is that the cross-reference does not have to be textual any more. This makes it possible to thread a search not only through a series of references, but also through illustrative artifacts on a page. For instance, in a litigation support activity, a stain on a photograph can be directly shown by evoking its exhibit number.

These capabilities make the use of optical disks a strong contender in the information retrieval arena.

---

## 2.2 IMAGING SYSTEMS AS INFORMATION TRANSFER AGENTS

---

In less than five years, the development of document imaging systems seems to have followed two parallel, although contradictory, tracks.

1. In terms of size, document imaging systems have grown from stand-alone "electronic filing cabinets" to installations supporting mission-critical applications.
2. In terms of relative importance in system design, their role has often been reduced from providing dedicated solutions to specific information control and processing problems to merely serving as efficient front-end data capture subsystems.

These tracks reflect two complementary perspectives of the same information delivery objective, which is to make information relevant to the conduct of business and available to the end users in a complete, accurate, timely, and cost-effective manner. In that context, document imaging ceases to be a miracle technology, capable of bringing instant cost reductions and productivity increases regardless of applications. On the contrary, it is increasingly evident that it is a very powerful tool, which, when fully and intelligently deployed, will enhance a business process. As a matter of fact, it can be shown that, when it is properly integrated in an organization's workflow, it is an essential component and sometimes a critical one in the optimization of an organization's business process.

Given the fact that an estimated 80 to 90 percent of an organization's information has its source in a paper medium, document imaging is the fastest and least expensive method for inserting that information into an electronic medium. Consequently, that technology provides the needed front end to an integrated electronic data interchange (EDI) process.

---

## 2.3 INDUSTRY TRENDS

---

Current document imaging systems are the result of several converging technological trends. As such, they can incorporate the features of different legacies, ranging from records management to data processing and telecommunication.

Early document imaging systems essentially were implemented as electronic replacements for paper files. At that time, the only practical path to electronic document sharing was through a mainframe. The economics of computing, on the other hand, caused a segmentation of the imaging industry between stand-alone systems for simple storage and retrieval tasks and mainframe-based installations, which could deliver the processing power needed for more complex applications.

As mentioned earlier, the sustained increase in micro- and minicomputer power and the decrease in hardware and software development costs have contributed to



cost-effective imaging applications outside the mainframe world. At the same time, progress in network technologies has made it possible to establish powerful information systems that do not need to rely on centralized processing.

Early document imaging systems can be classified in two groups:

1. Mainframe- (and to some extent mini- and midi-) based systems, where all resources including scanners, printers, and workstations are essentially peripherals connected to the mainframe, which holds the database and allows powerful multitasking functionalities. In these early systems, all the information processing power was concentrated in the mainframe.
2. Personal computer- (PC) based systems, which could be seen as a scaled-down version of the mainframe model, but with reduced capabilities, due to the limited processing power of their operating systems.

The structure of these two groups has traditionally been the result of technical as well as organizational politics and economics. Consequently, early on there was a wide gap between the computation-intensive requirements of workgroup applications, which only mainframes could support, and those of simple tasks such as word processing and spreadsheets, which easily could be accommodated at the PC level.

The competitive pressures of the hardware environment and the resulting production of increasingly powerful processors have built a bridge across that gap. The increase in processing power and storage capacity directly available to a PC has made it possible for microprocessors to begin supporting applications that only a few years ago required a mainframe, or at least a substantial mini or midi. In particular, the early selection criteria on what application could be efficiently designed for what platform (which were based on the maximum length of the basic instructions that could be processed by a given platform: 8 bits in the early PCs, 32 bits and over for minicomputers and mainframes) have somewhat disappeared with the advent of modern PCs capable of handling, sometimes with some "sleight-of-digit," 32-bit instructions. In addition, more efficient internal architecture and so-called cache memories can be used to smooth out inherent performance problems of lower-end systems. Now it is possible to port many mainframe-based, high-end applications to less expensive installations, thus enabling the implementation on local workstations of increasingly sophisticated and "power-hungry" applications. In addition, the increase of local processing power brought about the development of multitasking—or at least concurrent applications—a capability that previously had been available solely on mainframe installations.

At the same time, the development of network technologies has made the distribution of processing power along data utilization rather than resource lines possible, thus reducing the need to retain all information in a central mainframe. A major benefit of this development has been the ability to assign processors to specific tasks, such as file storage and printing, allowing for device sharing, thus making more processing power available to a user workstation.

These technological changes also have effected new challenges in imaging systems design, where the trade-offs needed to optimize performance involve the

choice of architecture, underlying processing power of the supporting platform, and the relative importance of required functionalities for a given overall information delivery system.

While most of the earlier imaging systems implementations were essentially stand-alone installations, present systems are almost exclusively used in a network environment and may or may not be integrated in an overall data and information processing environment. An outgrowth of this is the adoption of a wide range of document imaging applications, where these technologies can be used as a mainstream rather than as so-called orphan technologies, to optimize operational and technical solutions for identified business process requirements.

The continued progress in workgroup computing also is presenting a new challenge to the imaging system designer. The proliferation of so-called personal document imaging systems (PDoc), with local mass storage capabilities, suddenly means imaging can be introduced at a much lower level of granularity in the overall information system, forcing at least a review, if not a reevaluation, of the architecture and its image delivery capabilities.

Business issues in the imaging industry are further complicating systems selection decisions: The accelerated introduction of increasingly powerful platforms and increasingly sophisticated software tools, as well as a shrinking vendor base for differentiated products and services, is a problem for decision makers.

---

## 2.4 THE ROLE OF IMAGING IN INFORMATION TRANSFER

---

Regardless of their capabilities, imaging systems are expected to preserve the integrity of the original documents. Therefore, the role of imaging, per se, is limited to accurate recording, storage, and display of documents in a computer-processible format. As such, it could have remained an electronic replacement for microforms. The capability to extract electronic data from captured images and include that data in a larger information base gives image-based systems a much larger business value than stand-alone systems. But one must be fully aware that the more sophisticated the information system, the smaller the contribution of the imaging system itself to the overall information transfer process.

In fact, the decreasing costs of imaging hardware and the modularization of the software that controls imaging functions are generating user preference for off-the-shelf, "plug-and-play" solutions. This trend may give the system designer a false sense of simplicity in system implementation, which is reinforced by the vendor's promotional literature. This may actually have two different effects on imaging systems in particular and information systems in general:

1. The first, which is beneficial in the short term, relates to a reduction in design, development, and implementation costs of technical solutions.



2. The other, which may have long-term negative consequences, is the potential for "sclerosis" in systems design where only standard solutions, based on a limited number of products, are preferred to optimized implementations.

## 2.5 VARIOUS TYPES OF IMAGING SYSTEMS

Imaging systems fall into three main categories, corresponding to the role of the imaging function in the overall system. While there is a significant amount of overlap between the categories, that role cannot be dissociated from functions already supported by existing data processing installations. Specifically, transaction systems may use imaging to reduce the cost of data entry; business decision systems may use imaging to keep both documents and data about them organized; finally, archival systems will be used to reduce the cost of document retrieval.

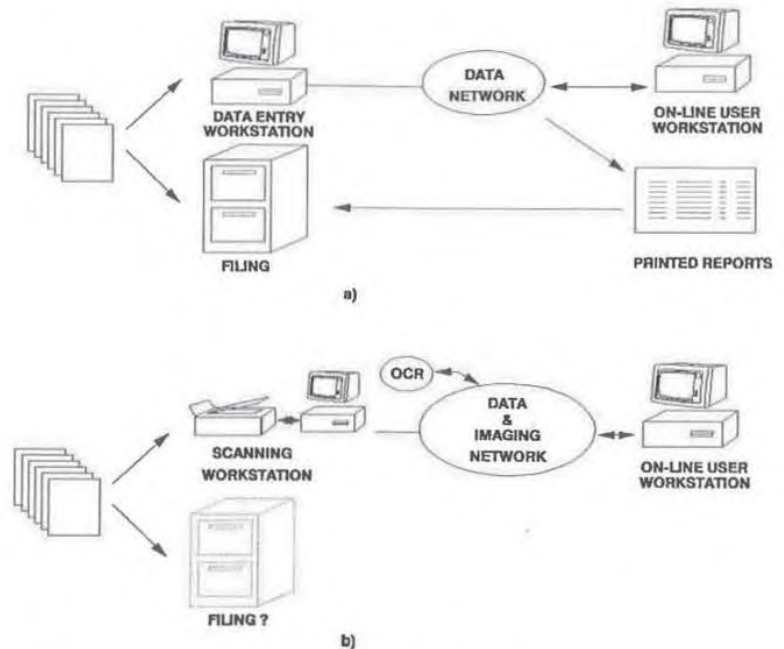
### Front-End Applications

In traditional data processing systems (see Figure 2.1a), data from documents are captured on data entry workstations and, after processing, are made available to end users through a network. At the same time, supporting documents are filed and archived for future reference according to standard records management practices, which involve manual filing in the office followed by off-site archival storage.

In imaging systems (see Figure 2.1b), documents are captured on a scanning workstation, and the data they contain are either extracted through keyboarding from the displayed image or, if the document is amenable to automatic data extraction, extracted through optical character recognition or similar techniques. After processing, both data and supporting image are then made available to the end user through an image-capable network. Because both image and information contents are then available in an electronic form, there is usually no operational requirement to retain the original paper, even though there may be other considerations, legal or historic, which may require that originals be archived. However, such storage need not be done on-site, thus reducing the cost of storage in valuable office space.

### Management Applications

Once documents and their contents are in an electronic form, they can be indexed and accessed in an orderly way, allowing for their systematic processing without loss of information. An increasingly important management application of imaging systems is associated with the management of the workflow throughout an organization, in which digitized documents are automatically moved electronically from workstation to workstation for processing according to established procedures.



**Figure 2.1** Using imaging as a front-end application

This ability to prescribe and control document progress electronically throughout an organization facilitates changes in document routing to accommodate different operational or business conditions, such as automatic escalation of unresolved issues or workload balancing. This new ability to manage the organizational workflow provides an opportunity to review an entire business process, which, in some cases, can lead to a complete redesign of existing data processing applications, with reported system maintenance cost reduction of better than 50 percent. This redesign process is discussed in more detail in Chapter 7.

### Archiving Applications

Archiving imaging systems are found in situations where large amounts of documents are to be kept for a long period of time, where retrieval time is an essential component of an organization cost, or when speed of access is essential to organizational purpose. For instance, organizations involved in a regulated business may be required to collect large volumes of documents; at the same time, it may be critical to the organization that it be able to locate a small number of documents on a moment's notice. One example of such a situation is in the chemical industry, where companies are required to keep track of all the chemicals and



compounds utilized and of the hazards associated with their handling and storage for use in emergency situations. Another example—often seen as a management application, but which is more akin to an archival system application—is in customer services areas, where the ability to settle a question while in direct communication with an inquiring customer must carry over long-term retrievability for audit purposes.

## 2.6 TECHNOLOGICAL CONCERNS

The penetration of imaging systems into the information processing industry, as well as the resulting challenges and opportunities, is the outcome of various factors. The rapid pace of change of computer equipment, software, and services, both in cost and capabilities, continuously modifies the technological landscape. At the same time, these changes create serious difficulties in planning and selecting an optimal technical solution to a given information transfer requirement. For instance, previously discrete functionalities to manage information flow, which formerly would have been designed individually, are constantly being integrated into increasingly powerful devices and software packages. This creates a “pause-or-play” dilemma, well known to the designer of computer systems. One must choose between waiting for cheaper solutions, and absorb a continuing drain on organizational budgets, or implement existing and presumably more expensive options, and reap immediate financial benefits. The answer to this is that the best time for a new system is probably the present time, because no doubt there will always be more powerful systems with more features in the future. Waiting for them, however, will exact a higher toll on organizational resources than need be. A further twist in the decision-making process is the fact that, because imaging technology is moving so fast, there will be little opportunity to gather information on prior experience.

Simultaneously, the coinciding decline in equipment costs and increase in processing power make the selection of “the best” technical solutions to a given problem an increasingly difficult task. More often than not, the designer will have to settle for an optimized “better” technical solution, rather than the one and only best configuration. The following list defines some of the issues facing a system designer.

- Intelligent workstations are becoming more proficient at performing local data processing tasks that were previously done on central mainframes. The question is whether to continue supporting a mainframe operation or adopt a “workgroup computing” philosophy, in which data processing is done at the workgroup level and then circulated in a “digested” form to all users of the overall information system.
- Low-cost multitasking environments enable intelligent workstations to perform local parallel processing of tasks, a capability that had been confined to mainframe installations. This raises issues of efficiency, distinct from the mainframe-workstation debate. One has to decide between concurrent task



performance on the same workstation, as opposed to "assembly-line" processing on separate workstations, as in traditional situations.

- Specialized workstations dedicated to communication tasks allow the proliferation of distributed processing applications. This creates new design opportunities for system architecture—combining data capture, storage, processing, dissemination, retrieval, and use—that best support business requirements in terms of performance, cost, and competitive advantage.
- Inexpensive personal document imaging systems (PDocs) are replacing, in some cases, centralized image scanning and storage, trading off central control for desktop flexibility in imaging system design.

Economic considerations provide their own set of constraints: Business conditions may not be conducive to the adoption of a fully reengineered information system. In fact, earlier investments in hardware and software may have to be taken into consideration. These constraints may change an overall project from delivering an optimal system to simply providing a temporary solution "until things get better."

Political and individual factors further cloud the design of a document imaging system. Availability of a "technical champion"—assignment of project responsibility to a given organizational unit, personalities involved, and even the training and experience of the project team—will have a strong influence on the system design and the success of its implementation. For instance, the design of an image-based accounts payable system will likely be different if the project is led by the financial area or by the MIS area. The finance people may lean toward the use of PC-based solutions, with high emphasis on controls; on the other hand, the MIS department will likely be partial to a mainframe-supported design, with emphasis on power. Systems design may also be affected by the fact that self-taught computer professionals usually are more comfortable with "off-the-shelf" packages than with formal programming and analysis.

Fortunately, the current level of technology may provide the means to get around the risks associated with the "wrong" technical solutions. With the accelerated adoption of standards for information interchange and software design, one can hope that, sometime in the relatively near future, technical options compliant with standards will be interchangeable at minimum costs. For instance, if an increase in functionality requirements tasks an existing system, one may need only to change the processing platform, without having to rewrite the application programs. The early use of process simulation can already reduce the risks of such migration by identifying optimized alternatives.

---

## 2.7 IMPLICATIONS FOR SYSTEMS ARCHITECTURE

---

In this climate of constant change, the key to long-term stability of a given design will be the development of an approach that closely shadows the business needs of the organization. This can be achieved by a modular architecture, providing a value-added service consistent with the contribution of an organizational unit to



the overall information flow. For instance, in an accounts payable application, the value added by imaging technologies may be limited to the receipt and verification of payment documents and for audits. There is no image-related value added for the actual processing of invoices. Consequently, the business rationale for changing data terminals to imaging terminals must be limited to those workstations assigned to material transaction verification, such as customer service.

Whereas the existing architecture may have had to deal with only a limited, well-defined set of devices and peripherals, the new architecture will have to support a wide variety of devices with increased throughput requirements and will need to be agile enough to adapt to rapid changes in business requirements.

---

## 2.8 A MULTILEVEL NETWORK FUNCTIONAL ARCHITECTURE

---

The design of an agile imaging system requires the resolution of a series of issues and problems unknown to the conventional machine-readable world. In particular, it will be necessary to know not only the volume of data traffic, but also to understand the nature and chronology of the various information processes involved.

For instance, not all imaging applications will involve a one-to-one relationship between document queries and image retrieval. In forms applications, where data are character strings in predetermined zones on the documents, the ratio between raster image size and forms content in machine-readable form is often 100 to 1 and higher. In many of these applications, there may be a 10 to 1 probability that the machine-readable data will be sufficient to answer an inquiry. Everything else being equal, the combination of these two factors will, without the benefit of a clear perception of business and retrieval requirements, yield designs with vastly different levels of performance. Simple "image enabling," in which low-end imaging systems are added to existing networks, is counterproductive when image retrieval, for display or printing, is more than even a few percent of the total number of queries. Incorporating high-end systems to an existing network, under the assumption that all documents need to be retrieved as images, may result in overdesign, at least in the short run.

### Functional Factors Influencing Imaging Network Design

Efficient design requires the combination and optimization of various requirements, such as workflow and architecture definition requirements, indexing, database software, and a graphical user interface.

#### *Workflow Requirements*

Because imaging offers the ability to integrate documents and data about them in a uniform architecture, review and restructuring of the way information should flow

throughout an organization will often be required. This in turn may result in a total redesign of an organization's underlying business process and supporting workflow.

### ***Architecture Definition Requirements***

The system architecture for networked imaging applications must be capable of supporting fast transfer of the large data packets representing digitized documents. The selected architecture probably will be based on a combination of the three common architecture models known as centralized, client/server, and peer-to-peer, which will reflect the kind and volume of data traffic.

To be complete, the architecture selection process must also involve the choice of physical and logical topology; protocols; operating systems; specification of connecting devices such as repeaters, bridges, routers, and gateways; cabling media; workstations; peripherals; mass storage; file/server platforms; and utility software.

The choice of architecture and supporting hardware and software will determine the maximum capabilities of an imaging system.

### ***Indexing***

Corresponding to the choice of the system architecture, which represents the "technical side" of an imaging network, indexing represents the "intellectual" side of the information transfer process. The word, by itself, has many meanings. In imaging systems, indexing refers to the machine-readable data stream assigned to a document for identification, storage, and retrieval purposes. One must realize that several indexing schemes coexist in an imaging system. Some are simple and transparent to the user, such as image location on a disk; others, such as document descriptors, are user-defined; still others, such as full text, are user-driven but automatically generated.

Because images are, in themselves, meaningless binary streams, the machine-readable data stream must be generated outside the image, either through keyboarding or by automatic data extraction.

The cost of indexing is, with document capture, the most expensive itemizable part of the operation of an imaging system.

### ***Database Software***

The database software used to fetch and display or print a document over the network links systems retrieval to network performance. Like indexing, many database software programs coexist on a network. Some, such as the database that manages the distribution of documents on mass storage devices, are internal to the system and transparent to the user and optimized to speed up the physical access



to the stored documents. Others, such as the image database retrieval software proper, are optimized to speed up the identification of relevant documents.

There are different types of image database software, the most popular meeting Structured Query Language (SQL) standards. However, these software may not provide efficient information retrieval in an imaging environment. It must be understood that retrieval in imaging networks involves a sequence of operations much more complex than for regular data, including storage and decompression delays. Consequently, it may be difficult to identify loss of degradation of performance as resulting from an overtaxed database or other causes.

### **Graphical User Interface**

With the practically universal acceptance of graphical user interface in most workstations, it is necessary that the requirements for imaging be compatible and coexist with the other programs on a user workstation. Because both imaging and other user applications will compete for computer and memory resources, their choice must not be taken for granted.

Under those conditions, efficient imaging network architectures will usually require the use of a multinet arrangement, composed of several levels, involving different subnetworks, each subnetwork being optimized for the performance of specific tasks.

### **The Different Levels**

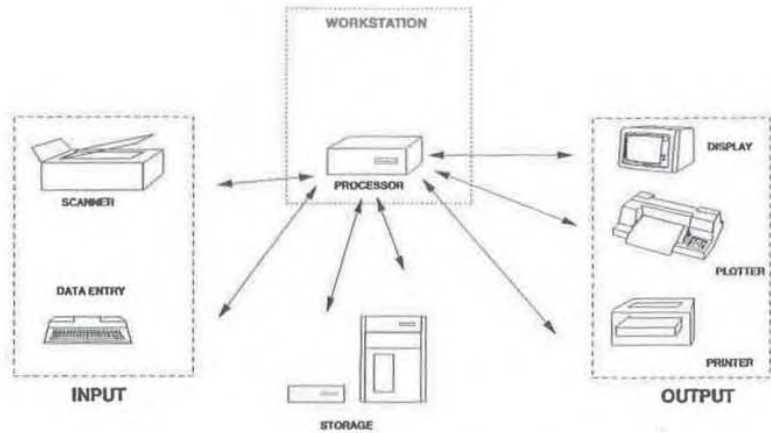
Efficient imaging network architectures may be composed of several levels, involving different subnetworks, with each subnetwork optimized for the performance of specific tasks.

There are several accepted models that provide a formal description of the processes occurring in a network. The generally accepted model is the so-called OSI model, which assigns all data functionalities found in a network to one of seven hierarchical levels (see Chapter 3). IBM System Network Architecture (SNA) adds to that seven-layer model specification for network nodes and links, and for the actual interface and network connections.

From a functional standpoint, a networked imaging system can be described in terms of four different functions: user interface, imaging and other user applications, information capture and control, and document storage and retrieval. These functions are supported in several places in a network. From the user point of view, they can be located in one of three levels: desktop, processor, or network.

#### **Desktop Level**

The desktop is the combination of the display screen, keyboard, and pointing devices that provide a functional interface between the user and the computer. It



**Figure 2.2** Basic components of an imaging system

can be stand-alone or connected to a network. From a user point of view, as diagrammed in Figure 2.2, the configuration of the workstation is of no importance as long as it allows retrieval and processing of information in a timely, complete, easily readable, and accurate manner.

The minimum functionalities that must be provided at the desktop level include image-specific tasks (e.g., inquiry, retrieval, display, printing, storage) and non-image-specific tasks, which permit access to all the applications that are performed through that desktop.

### **Processor Level**

The processor level is the functional interface between the desktop and the rest of the network system. The processor is designed to perform the tasks it can do locally, to forward through the network the tasks that must be performed by other processors, and to know the difference between the two.

While the desktop level is relatively easy to configure functionally, the configuration of the processor level requires a thorough understanding of the applications that must be run from the desktop so that the workload can be allocated between the user processor and other devices on the network. For instance, scanning can be performed in a distributed way, either locally or remotely, or centrally. Understanding the use of the digitized documents will facilitate the decision as to how the process should be carried out, and will help to achieve the optimization of the processor configuration.



### **Network Level**

The network level corresponds to all the functionalities that are not processed locally. While the two other levels can conceivably, from a user perspective, be configured independently, the network level usually requires the contribution of network specialists to design and optimize this third level of functionalities.

---

## **2.9 IMPLICATIONS FOR USER INTERFACES**

---

Because the desktop is the functional interface between the user and the computer system, it is also a bottleneck in the information transfer process. Graphical user interfaces and multimedia user interfaces (MUIs) are essential to minimize the impact of that bottleneck on desktop efficiency. Three functionalities must be supported by the interface: image display, commands access, and controls and management.

### **Image Display**

The system display must meet a wide range of requirements, from high resolution for the scanned images, to the lower resolution for display of icons and commands in a windowed environment. Because each requirement corresponds to different demands on supporting hardware and software, it is important that the system design be compatible with the capabilities of the display. For instance, some document-intensive applications may require simultaneous display of a large number of scanned images, together with geometric graphics and machine-generated text. The GUI must be able to integrate all these requirements (multiple raster, vector, and character-based data) into a single display.

### **Commands Access**

The issue here is the integration of user interaction with the system without interfering with ongoing processing of applications. Each GUI provides tools for exercising such interactions. However, the ways by which GUIs place calls to the system vary from system to system, and therefore the speed at which a given command is executed from the desktop depends on the particular GUI.

### **Controls and Management**

This series of tasks corresponds to the display of system messages and programmed status reports. Depending on the nature of the messages, they may be character-

based text or embedded with graphics to reflect a particular condition or attract attention.

GUIs require intensive computing resources. As such, they compete with the resource requirements of the imaging application itself. Consequently, the design of a networked imaging system must be done with due consideration for the potential conflicts in the local processor for access to the desktop.

---

## 2.10 STRUCTURE OF AN OPTICAL DISK SYSTEM

---

This section focuses on the physical components of an imaging system. As such, there will be some overlap with concepts already covered. However, the description that follows will now be device-bound, rather than information-bound. Consequently, the discussion of a given device, for example, a workstation performing several different functions, such as display and processing—functions kept separate in the past—is not a repetition of previous material but a reflection of the complexity of networked imaging systems.

All document imaging systems are an assembly of basic components, namely scanners, processors, storage, displays, and printers connected by communication links (see Figure 2.2). Basic imaging functions are as follows.

### Document Capture

Documents are captured (input) into the system through a scanner that provides an electronic representation of the optical image of a document, which may be on paper or microform. The electronic representation of the image is initially in an analog form, which is then translated into a digital signal.

### Processing

Processing incorporates several different functions, which cover acquisition, indexing, storage, retrieval, display, printing, and communication of images. Processing tasks are performed on one or several computers, which are loaded with software and processing boards specifically designed to handle quickly the large amount of data contained in an electronic image.

Most often, indexing, storage allocation, information retrieval, communication, and system management are handled in software. Hardware, on the other hand, usually handles computation-intensive operations, including data compression and decompression, format translation, image manipulation, and device interface between systems components. However, the rapid adoption by the marketplace of high-performance processors is making it possible to design systems in which tasks that were initially done in hardware, such as decompression and format conversion, can be executed in software at a lower cost.



## **Storage**

The storage function is usually performed as a combination of magnetic (i.e., erasable) and optical (erasable and nonerasable) memories, where the magnetic memory is used as a scratch pad during the image capture process. It is also used to store indexes to the image locations, as these need to be updated on a constant basis. The optical disk is used to store the permanent images. In systems using nonerasable disks, updating of an image is done by recording a new image on the optical disk, and amending the index on the magnetic disks to reflect the change. Optical disks can be kept in stand-alone drives or in what is called a jukebox, which is a device equipped with automated disk changers, thus allowing continued access to data. Depending on the number of workstations on the system and the frequency of access, a jukebox might include several drives working in tandem to meet needed response time.

## **Display**

Display and printing devices are peripherals that make it possible to visualize and generate a physical representation of the digitized document. Later in the book, the major issues raised by the various models of GUIs, which ensure human-machine communication, are discussed. Because of the differences and limitations of the physical processes involved, there can be significant variations in overall performance and quality between displayed images and printed hard copies, as well as in outputs provided by different equipment.

In special applications, such as publishing, printing may be done on film or on other optical disks. A discussion of the application to microform production is provided in Chapter 16.

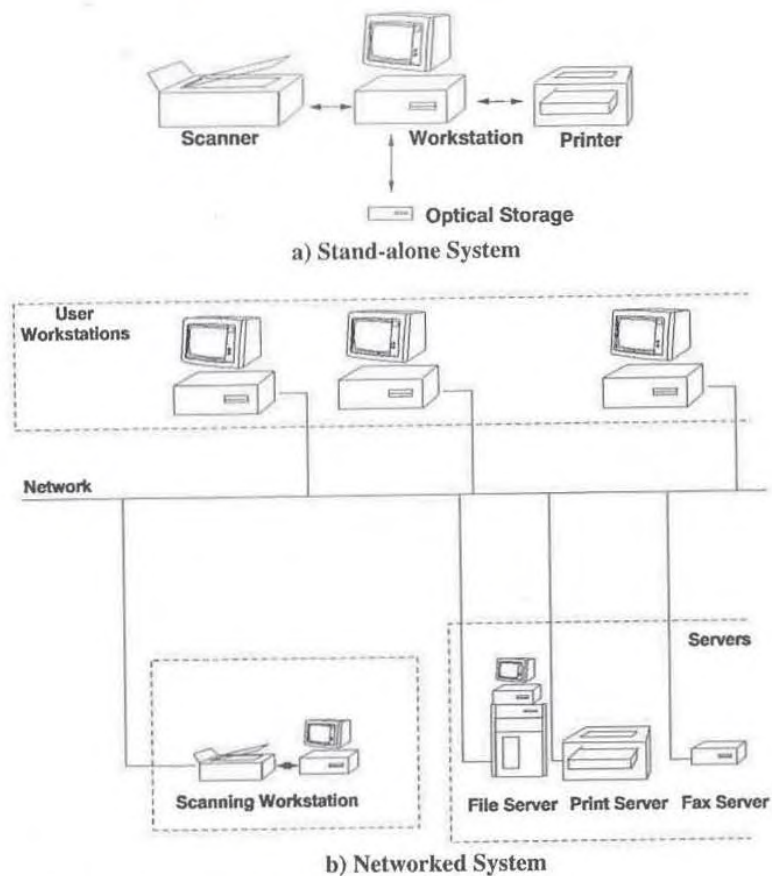
## **Communication**

Communication tasks cover two different functions: One deals with the transfer of data within a given processor and between the processor and its associated peripherals; the other relates to data transfer between several processors. These functions are best described in terms of configuration, and two main configurations can be adopted for document imaging systems (see Figure 2.3).

In a stand-alone configuration, the processor controls document capture, storage, retrieval, and printing. Communications are purely internal to the system. In that configuration, which may be practical for very low use, tasks of document storage and document retrieval cannot be performed at the same time.

In a networked configuration, document capture, processing, and storage can be separated from retrieval, and multiple users can have access to the same image base. Consequently, storage and retrieval tasks can occur concurrently.

In the context of this chapter, communication devices are defined as the devices that interface the optical disk system with the outside world. In simple (stand-alone) systems, communication devices are limited to the interface that allows correct



**Figure 2.3** Example of two typical configurations

transfer of information between the optical disk storage and the printer or display. In more elaborate systems, communication devices allow the interconnection (networking) of several units for the purpose of sharing resources or information. A detailed discussion of the issues involved in the design of a networked imaging system is provided in Part II.

## 2.11 EXAMPLES OF APPLICATIONS

Because optical disks make it possible to integrate data and images, they open processing opportunities that were previously impossible. For instance, in legal files applications, immediacy of image capture and availability of communication links allow attorneys to be immediately appraised of any change in an issue. And



because these functions are computer driven, access can be controlled under well-established computer security procedures. Furthermore, if the information is collected in a standard form, it is possible to design the application from which data on the form can be extracted directly. This eliminates the need to keyboard standard data such as names and filing data, while retaining the ability to capture non-machine-legible information such as signatures and similar information.

In regulatory and library applications, all documents can be integrated into a uniform display environment. In particular, word processing and imaged documents can be stored in the same database and retrieved under a consistent procedure.

With a detailed index, it becomes possible to eliminate filing, because the computer keeps track of the document automatically, and thus saves on clerical costs. The use of optical character recognition (OCR) algorithms can often simplify the task of indexing complex documents by extracting full text from imaged documents.

In applications related to transaction processing, the ability to provide multiple access to the same document can significantly shorten the overall processing time spent handling the document. If the system includes a jukebox, the index will automatically trigger the mounting of the appropriate disk and the performance of information retrieval.

Imaging can also be used as a front end to the preparation and use of interactive technical manuals, which then can be replicated on commercially available systems for dissemination.

---

## 2.12 DESIGN CONSIDERATIONS

---

Several considerations come into play when designing an optical disk application, and it is important that they be fully understood before a system is specified. In particular, because of the rapid evolution of technology, the relative importance of these considerations in design might vary within a short period of time. For instance, because hardware price performance ratio is increasing, the cost of hardware replacement is becoming a small factor in the overall system cost, forcing the designer to focus attention on design and conversion costs.

The first consideration is to ensure that information stored on optical disks is retrievable in a complete, accurate, and timely fashion. Because optical disks are not human-readable, it is essential that the procedures for collecting, recording, processing, storing, retrieving, displaying, and printing information guarantee the integrity of the information, an issue often referred to as the legality of optical disks. In most cases, if the procedures are well defined, and image capture is done in the due course of business, optical disks can conceivably replace microforms as a storage system, and the original documents can be discarded after recording. In other cases, the application might be such that optical disk systems serve an operational role, with the original documents kept for evidentiary purposes.

Other considerations are related to the need for all systems components to be compatible among themselves, and that they remain that way for the foreseeable future. Because of the lack of standards in the industry, compatibility might require



selection of specialized interfaces to allow noncompatible machines to operate together and still maintain good performance levels. Of more importance is the concern that a particular component, or even a full system, will still be supported in a not too distant future. This consideration is usually handled through a cost-benefit analysis, where the overall system cost is established to provide a quick payback period. Because a conversion from one digital system to another can be done automatically, the cost of conversion is essentially a one-time charge. Under that condition, should there be a need to change the system, the replacement cost would not offset initial benefits.

In designing the overall system, answers to questions as to its size and the interconnection of its various functions will determine its overall architecture. The application of records management techniques that keep files to a minimum without sacrificing operational efficiency, in concert with a full understanding of the business requirements of the applications being supported, optimize the design for the current needs, and at the same time provide the opportunity to prepare for upgrades by embedding access points for future developments. Of particular importance is an assessment of growth requirements, so that the indexing system can absorb additional storage, such as linkage of additional jukeboxes, without loss in response time and other performance elements.

Related to the consideration of systems architecture is the issue of networking. While records management and workload studies can provide the data indicating the number of workstations to be used on the system, the question will be how to connect them. Of particular interest will be whether all stations should be equipped with a full complement of scanners, high-resolution displays, and printers, or whether device sharing will be acceptable. Attention will also have to be given to interaction with different systems. For instance, a personnel system might need the information that is in a payroll system, and there might be an operational value to having both systems share a common database. The same situation might involve a purchasing and accounts payable department. Different requirements will yield different types of networks, from fully centralized to fully distributed.

A last consideration before completing the design and selecting the hardware relates to conversion to the new system. In applications such as accounts payable, one might choose a day-forward system, in which only new files are input into the system, using the argument that there is a low probability that older files will be accessed, and therefore there will be no financial benefit in bringing them into the system. In other situations, such as personnel, it is imperative to design retrospective systems, where all files are converted. Not doing so would prevent file integrity.

All these considerations directly influence the design of the database supporting the optical disk system and the nature of the indexing scheme to be used. For instance, if all the information/documents stored are composed of forms, the database can be built on a flat-file design, with simple indexing. If, on the other hand, full-text documents are part of the document base, the database might require more complex features, including a relational structure; the indexing system might involve the capability of performing full-text searching. If conversion is involved, plans will have to be made as to the nature and compatibility of the conversion equipment and its integration with the equipment used for continued operation.



## 2.13 IMPLEMENTING AN OPTICAL DISK SYSTEM

The implementation of an optical disk system requires a well-orchestrated succession of steps, beginning with collecting baseline data. These data will enable the designer to size the project, to ensure that the design meets the expected level of performance. For instance, an initial study will be necessary to assess workload data so that the number of workstations is adequate and that the size of the processor is sufficient to handle the volume of transactions within acceptable response time. This might be done through a survey to identify core needs or those requirements that are essential to acceptable performance. An analysis of the current information flow will make it possible to determine the critical information path or the relationship between various functions, together with the identification of retrieval patterns and critical response times. The use of models and simulation can greatly improve the ability to specify the new system.

Once the baseline data have been collected and formally analyzed in a requirements analysis, the actual design can begin, starting with system and structure definition, and including functional requirements, hardware selection, and performance specifications. Of critical importance will be the design of a database structure consistent with the business needs of the organization and the selection of the database engine that will support it. Powerful design tools for efficient operation will include modeling and simulation.

The overall technical solution should then be submitted to various tests of acceptance, including budgetary value and admissibility of procurement procedures. Of importance to the design will be considerations of the life cycle of the proposed solution and whether the implementation will be conducted with in-house staff or through outside contractors.

Once the design has been finalized and accepted, its implementation will require installation planning, testing, and training, often with conversion serving as a prototype of operation. In the process, several issues will need to be resolved, including retraining or restaffing for the new system and the learning curve involved in implementing a new system. As found in the data processing field, it will take several months before trained users are able to articulate any requirements for enhancements, thus requiring system support beyond the initial startup period.

One of the most significant elements in the implementation of imaging systems is the cost of conversion and/or indexing for storage on optical disks. These costs can often be offset by savings from the elimination of filing tasks. In certain cases such as legal applications, such costs, which are already part of the legal file operating expenses in a manual environment, can be reduced through the use of OCR, which, in many cases, allows automatic extraction from the document.

Once the documents have been indexed, they can be arranged and displayed or printed at will, without having to physically handle the files. This gives, in a litigation support application for example, the ability to create, on demand, packages of documents that can be viewed either in a hypermedia/window fashion or distributed on paper, which need not be kept after the case has been settled.

Other facilities derived from the electronic image include the ability to directly receive and store facsimile transmissions as an integral part of the files and the ability to have several people working on the same file at various locations without risking a loss of file integrity.

---

## 2.14 CONCLUSION

---

Electronic imaging is a powerful tool for information retrieval and an efficient front end to an integrated information system. With a proper design, imaging systems can present significant cost advantage and additional benefits over manual and microfilm-based systems.

As indicated earlier, however, the adoption of a system should not be a trivial decision. Aside from the fact that implementation costs may be considerable, imaging systems represent a new way of doing business, and their acceptance may be an irreversible decision. Therefore, it may be beneficial to an organization to secure, either internally or through outside counsel, the necessary know-how to ensure success in implementation.



---

# 9

---

## Image-LAN Enabling and Integrating LAN-Based Imaging Systems with Existing Mainframe Applications

---

*Peter Dolshun*

*Sigma Imaging Systems*

*Marc R. D'Alleyrand, Ph.D.*

---

### 9.1 INTRODUCTION

---

In most systems, the issues of data ownership, usage, and custody are at the heart of information flow strategy decisions. The resolution of these issues supplies the rationale for funding existing operations, the cost justification of new projects, and, more importantly perhaps, for the political support to achieve change. In that respect, workflow redesign focuses not only on operational perspectives, such as transaction volumes and database integrity, but also on the identification of access rights. In particular, consolidation of tasks is one of the many processes where the issue of data ownership must be well understood.

For instance, does an accounting department have "ownership" of the data that go into a general ledger because that department is the only one empowered to make entries in accounting records? Conversely, does this department then only have custody of the supporting documents because they do not generate them? And does a budgeting department have ownership of reports that they prepared from account-

ing data, when, through network and dynamic data exchange linkage between programs, the underlying figures can be automatically adjusted and replaced in the report itself, possibly not only invalidating earlier analysis, but creating discrepancies between data and conclusions without the knowledge of the reporting party?

In earlier data systems, such issues could be easily resolved by providing different access rights to different users, assuming, for instance, that "writing" privileges are equivalent to ownership. This simple approach might resolve the ownership issue for initial entries, but it does not settle the case of derivative documents, generated as composites from different sources. From that standpoint, image enabling an existing network is a complex operation, because, at the same time that the system brings together data and images, it merges the two hierarchies of data ownership based on automated and manual systems.

The difficulties in assigning responsibilities for data ownership and access are exacerbated when dealing with the integration of a mainframe operation with an imaging network. Here, however, even though the technical obstacles may be significant, they are predictable and well defined. In fact, they may be considered minor when compared with the issue of potential loss of total control over database and system architecture resulting from the integration of the two systems.

This section describes the issues to be addressed in managing the integration of a distributed LAN-based imaging system with a mainframe-based transaction processing application. Making the proper selection of an implementation approach has deep-rooted strategic implications that can save (or lose) a substantial investment in scarce capital, personnel, and other corporate resources already invested in the mainframe environment. Doubtless, a LAN-based imaging integration *can* yield a more flexible, better performing, and more cost-effective operation than a centralized mainframe imaging application, but final benefits depend on the quality of the integration effort.

Mainframe application design and maintenance methodologies are well known; however, the changes necessary to integrate such applications with a LAN-based, client/server application may prove to be somewhat of a challenge. One of the major issues that will need to be decided by the various stakeholders is how much of the mainframe responsibilities—both data and processing—should (or could) be relinquished to a LAN. Early on the answer could have remained purely a technical one (following the argument for instance that some applications could simply not be "trusted" to a LAN). The changes in hardware, software, and database technologies are making such an answer more dependent on political and financial criteria. For instance, the introduction of Windows 95, an operating system rather than an environment, is eliminating some of the major limitations of windowed applications by allowing true multitasking, an impossible capability in Windows 3.1. However, even with its enhanced capabilities, Windows 95 will not likely be able to support all applications that have been initially designed for a mainframe, and, even if it could, the cost of migration to a full LAN-based system could rarely be justified. Nevertheless, some applications may be better handled at the local level, and, consequently, there will be cases where a partial migration may be justified. One example of this will be when image enablement will require reliable access to data stored on the mainframe. That reliability will be obtained through integration, and



this chapter will focus on some of the details that must be attended to in order to facilitate the planning and implementation of such an integration.

## 9.2 CURRENT TRENDS AND CHALLENGES

Integrating existing mainframe systems with newer LAN and workstation technologies goes hand-in-hand with the migration of small mainframe-based applications toward less expensive and more user-friendly local area network departmental systems. LAN-based systems are already cost-effective, and reductions or outright freezes on new mainframe project budgets have further accelerated this movement. Thus, the interest in distributed LAN-based imaging systems and their associated integration tools is also extremely intense.

Technological problems still exist that prevent the conversion of very large-scale, high-volume transaction processing applications to a local area network environment. This will be a mainframe stronghold for a long time to come, especially in the largest corporate installations. Mature, sophisticated mainframe platforms still excel in areas of centralized control of massive data stores and high availability for hundreds or thousands of simultaneous users. So, while the desire is there to move in the direction of inexpensive, distributed PC platforms, the conversion of large applications that must support hundreds of users will take place only when the LAN environment is as reliable, secure, and as easily controlled as the mainframe side of the shop. But, more powerful hardware and more production-capable software (system operating software, network operating systems, and applications) are now available, and client/server software development techniques are maturing, so the trend to resist new mainframe development will continue in parallel with the growth of networked workstation systems.

In addition, risks associated with major changes to core business applications are a major deterrent to decentralization, unless there are overriding strategic or economic considerations. Furthermore, the political issues associated with loss of status in MIS departments as a result of potential large-scale downsizing, may create obstacles in securing investments for systems that may ultimately migrate from the mainframe.

Under these conditions, it may be risky to attempt the integration of imaging systems directly with host applications, as it may actually result in a destabilization of mission-critical transaction processing operations. The use of LAN-based imaging solutions does not present these difficulties and provides a perfect fit for this kind of environment. Existing applications do not have to be impacted at all in order to take advantage of today's latest document imaging systems technologies and the associated benefits. MIS management can take a low-risk strategy by providing a more flexible, transparent "add-on" solution with a LAN-based imaging system. By carefully defining the integration requirements, the existing base of mainframe applications (which may work perfectly well) requires very few changes.

For some companies with a heavy investment in mainframe applications, imaging implemented on workstations may also be their first exposure to a large-scale LAN implementation that does not require that they first redesign their



existing operation. In fact, the resulting installation of the LAN/WAN/workstation facilities allows the enterprise to initially take advantage of imaging capabilities and simultaneously open the door to new application possibilities. In this way the imaging system, which typically has a comparatively short payback period, justifies its existence and also paves the way for a more powerful future; and, through workflow redesign, it enhances the benefits of the new networked infrastructure.

### 9.3 IMPACT ON TRADITIONAL TRANSACTION PROCESSING OPERATIONS

Preserving existing systems during the transition to LAN-based imaging systems also benefits the user community, because it reduces the extent of workforce retraining. In many LAN-based image-enabling applications, the imaging system can coexist with the mainframe to support its simplest function: replacement of paper with the automatically controlled delivery of electronic document images.

Another advantage of the adoption of LAN-based imaging as opposed to mainframe installation is the fact that, typically, very few companies would install imaging systems throughout an organization without preliminary testing. Instead, small pilots are installed for evaluation and expanded if successful. This "scaling up" approach, which can be very expensive for centralized, mainframe-based systems, is much easier to put in place with a LAN installation, as it can be more easily scaled up to accommodate high production volumes by simply configuring additional servers and workstations as needed.

The impact of an imaging system installation will be felt throughout the enterprise; therefore, comprehensive, long-term planning is essential to prevent costly mistakes later on. For example, the lack of standardization across departmental implementations will create communications problems once the desire to access document images throughout the company is realized.

The software chosen must be able to handle protocol and data format conversions. Lack of a standard platform also leads to complex interoperability issues (which the enterprise must be willing to support). Certainly, interoperability between different systems is feasible, but it is advisable to keep the number of different platforms to a minimum to reduce the cost of maintaining heterogeneous installations.

It may take years to go from planning to implementation of a successful large-scale imaging operation; therefore, the challenges should not be underestimated. It is worthwhile to attempt a simplified initial implementation strategy in order to ensure early success. Most imaging systems tend to be analyzed and designed in terms of departmental business requirements, which makes sense, because individual departments are usually divided by lines of business, and each is held accountable for its productivity.

Each department processes a well-defined subset of the organization's total transaction workload. Under these conditions, departmental users are best able to assess how the introduction of a new imaging technology will affect their day-to-



day operations, and their needs must be the major guidelines for proper design and installation of the new systems. Positive user feedback early in the design and evaluation phase, as with any application, is crucial for initial acceptance.

---

## **9.4 MOVING FROM LOW- TO HIGH-VOLUME TRANSACTION PROCESSING**

---

Although the applications may be identical, low- and high-volume processing present structural differences, an important concept that is explained throughout this section.

### **Low-Volume Transaction Processing**

The type of operation that characterizes low-volume processing, such as would be found in a typical pilot installation, can be described in the following manner:

- The tasks are not mission critical.
- The system is often used in a departmental environment.
- The tasks performed relate to one type of work.
- These tasks are self-contained, and involve only a few people whose skills may be sufficient to run the application without additional training.
- The operation is well insulated from the rest of the organization and requires minimal external support.

These low-volume operations are often conceived as proof-of-concept rather than as part of an integrated design, and thus have few or no interface with the rest of the organization or dependencies on other external systems. Consequently, they may be difficult to expand or use elsewhere, as the system's functions may be too specialized or restrictive.

### **High-Volume Transaction Processing**

By contrast, high-volume transaction systems are designed to operate on a company-wide basis. Their characteristics include:

- Ability to simultaneously handle many complex and different transaction types.
- Workstations are geographically dispersed.
- Efficient use of imaging requires a complex image data storage hierarchy, which itself may be distributed.

- The system reaches almost all departments in the organization, regardless of responsibilities.
- Installation and maintenance of the enterprise system requires training and acquisition of new technical skills for a significant portion of the staff.
- The imaging system is based on a combination of core processing capabilities with standard interface and custom applications to accommodate departmental needs.
- The system is tightly integrated with existing internal and external installations.

### **Matching Imaging System Features with Mainframe Requirements**

Choosing the imaging system that matches a given mainframe-based installation depends on the nature of functions being supported. The recognition of transaction types and processing requirements is essential to that determination, because the performance of imaging applications as implemented will depend on constraints defined by the type of transaction being processed.

Some transactions are relatively simple (single-step data entry); others can be very complex (multiple steps, multiple departments completing different portions of the entire transactions at different times). Because of this, target operator processing rates may be different for each step in the workflow. The timing requirements of the mainframe system being integrated (batch versus on-line, multiple data capture and processing cycles, reporting and audit trail processing dependencies, batch and backup processing windows) are other major factors to consider. One size of imaging system may not fit all unless it is flexible enough to accommodate all of the organization's processing needs.

During the design and planning stages, transaction profiles should be examined and understood in terms of their image throughput needs. Different departments will have unique configuration requirements.

### **Examples of Transaction Types and Processing**

Imaging network requirements are dependent on the response time necessary to support the business process and the volume of data to be retained on line, as exemplified in Table 9.1. Typical examples are as follows:

- Applications requiring high-volume local server capacity, fast image retrieval, and display are characteristic of: customer service telephone inquiries; priority insurance claims processing (i.e., Medicare claims processing); single-step (simple), "heads-down" data entry.



Table 9.1 Volume and speed requirements of processes

	<i>Server Capacity</i>	<i>Retrieval Speed</i>	<i>Display Speed</i>
Telephone Inquiries	High	High	High
Insurance Claims	Medium	Medium (if under workflow control)	Medium
Correspondence	Low	Low	Low

- Applications requiring access to multiple mainframe applications, and for which image display performance is still very important, including: multistep (complex) data entry, health care claims processing, transaction correction, and reprocessing applications.
- Applications where fast image display is not critical, but workstation capacity and power are needed to access multiple mainframe and PC applications (such as spreadsheets and word processors) at the same time. These include: automobile insurance claim processing, insurance policy and contract underwriting, customer service correspondence processing, auditing.

At the same time, mainframe application access requirements will vary. Heads-down data entry applications typically require a single host session. Complex transactions may require access to many host applications and databases simultaneously in order to resolve a complex customer service inquiry, for example.

Since document images usually participate as information stores in many different kinds of organizational transactions, the new system must integrate and optimize interdepartmental workflows and routing capabilities, often under different retrieval operating conditions. The software should be configurable to minimize customizing programming to enable these features, because the underlying transactions, performance requirements, and organizations can change frequently, regardless of the complexity of the underlying business requirements.

## Measuring Results

In any image-enabling project, the measure of its success, let alone its justification, will be increase in user productivity, quantified for instance by operator's response time. The speed at which a randomly selected document image can be retrieved from anywhere in the organization will range from slow transaction response requirements (e.g., overnight audit project) to fast (e.g., customer service telephone inquiries requiring document verification).

Performance depends on a variety of factors, the most important of which is probably the design of document storage hierarchy, which is the automatic staging of work in anticipation of its retrieval by the operator in order to minimize response time. The following text discusses some of these aspects from the mainframe perspective.

### **Document Image Caching, Routing, and Staging in a High-Volume Environment**

Caching of work in progress is the most effective method to increase productivity. Assuming that a hierarchy of storage has been adopted, the main design issue becomes the definition of a document allocation scheme on local servers for departmental processing.

In most high-volume installations, documents to be processed are received in a central mail room and sorted by line of business. In image-enabled installations, the documents are then scanned at one or more locations (usually close to the mail room operations). A typical sequence of events looks like the one displayed in Table 9.2.

Staging document images from archival storage complements the caching of newly scanned work. Response time for archive retrieval (from optical disk jukeboxes), whether channel-attached to the mainframe or server-attached on the LAN, takes anywhere from a few seconds to several minutes; a platter mount for an off-site disk not resident in a jukebox may require an overnight delay. The way to maximize productivity is to minimize this response time through intelligent staging to hard disks on local servers. The definition and anticipation of the data are required to complete a transaction before the operator attempts to begin processing. This includes the document image, of course, but it also implies that the images are stored where the work is to be processed. How can this be done?

Given the wide variety of individual transaction types that an enterprise must execute, it is usually organized along lines of processing work and further subdivided by lines of business. For example, a health insurance carrier must divide the claims processing area into data entry, suspension processing, audit, and customer service departments. Within these, each department is split further by type of claim (medical, dental, prescription, etc.). This is usually done because each resulting division (dental data entry, for example) may require specialized operator skill sets for transaction completion. Mainframe application operators are usually aggregated locally along these lines.

For instance, distributed systems must be configured to contain locally all the data required to process specialized transactions. Since the departmental organization is well defined in terms of the type of work it performs, document images with the corresponding attributes can be placed electronically (under system control) where and when they are needed.

Once work is staged and processed, it must be deleted to provide space for new images. This may be done immediately after the transaction has been acknowledged as completed. For data entry operations, the decision is simple: Delete the document image immediately after keying and verification is completed; the imaging system can perform this automatically. The deletion decision can also depend upon the status of the transaction as determined by the mainframe application, in which case, the application must somehow communicate with the local departmental system and indicate that it is safe to delete a document. This is usually true of multistep transactions where the document image must stay resident until the mainframe application finalizes the unit of work. Productivity would be substantially slower if operators had to constantly mount optical disks for document



Table 9.2 Example of processing stream in an image-enabled operation

1. **Document Receipt:** Documents are received at the mail room. They are opened, date/time stamped, and sorted by line of business.
2. **Preparation:** Line of business documents are prepped (staples are removed, tears are repaired, etc.) and batched (counted and assigned control numbers).
3. **Document Capture:** Documents are processed with high-speed scanners, which create digitized images. In the LAN-based configuration, each scanner is controlled by its own CPU (a PC) with some hard disk storage to hold the scanned images. As each document is scanned, attributes can be assigned automatically by the scanner software. These attributes (an incremented control number, a date and time stamp, or a bar code found on the paper document) can be set manually by the operator, or automatically.
4. **Quality Control:** The scanner controllers, attached to the LAN with network adapters, can release the document images into the network and route them to an appropriate quality control workstation, which, in case of error, can delete the rejected document and send a message to the holder of the paper original to rescan it.
5. **Distribution:** Once the images have passed the quality control step, they are routed to their destination, determined by the attributes of each document. The imaging system software can be configured to always send one copy of each document to servers attached to optical storage devices (to accommodate archival requirements) and simultaneously send another copy to the appropriate departmental servers as work in progress.
6. **Mainframe Monitoring and Control of Image Placement:** The mainframe application does not need to know where the images are stored. Workflow software and the document database will maintain a directory of scanned documents, which may span several destination servers, established to hold specific subsets of work types known to the mainframe applications. The same workflow software will usually control the image placement throughout its life cycle and control the local server cache size, according to the number of devices attached and the volume of work being processed, thereby optimizing resources to ensure peak productivity.

image retrievals from the archive only. Local caching within departmental systems is the way to maximize throughput. Alternatively, a time-triggered process executing on the server can delete image objects based on a simplified criterion, such as last referenced date.

The management of this staging and deletion process is not as difficult as this description first seems. Normal transaction processing generally requires that the oldest work has priority and must be processed first. Many operations require that the day's receipts must be processed in a predetermined time interval in order to prevent a backlog.

---

## 9.5 PRODUCTIVITY FACTORS

---

Productivity increases associated with image-enabled mainframe applications depend on user acceptance of distributed imaging systems, as well as on technical parameters.

### Human Factors

Human factors begin with the commitment from the top down to the use of the new system to enhance long-term business objectives. Failure to do so prevents efficient communication between users and systems planners, and places obstacles in the path of management support. This may have disastrous consequences because, like an airplane taking off, there is a point-of-no-return in the development cycle of an image-enabled system.

### Technical Factors

Productivity is a prime user acceptance criterion driven not only by savings and profit motives, but also by industry regulations and legislative directives. The new imaging system must provide enhanced productivity in order to be considered a success. Transaction processing cycles must be at least as short and quick as existing applications. Given the large storage requirements for document images, this is an extremely complex technical challenge: how to rapidly present document images to terminal operators. This is accomplished through the use of operator workstations configured with enough memory and processing power to enable the caching of images at the workstations. Typically, the minimum configuration is a 386SX-class machine loaded with a minimum of 8MB RAM, but 486-class machines loaded with 16MB or more will become common as the cost of this hardware continues to decrease.

High-performance image presentation is much more difficult to support from a centralized mainframe application. Expensive mainframe DASD must be used, CPU cycles are much more costly, and network bandwidth is typically optimized to transmit short bursts of small 3270 data packets. By distributing the image-handling application over LAN servers and intelligent workstations, image services can be delivered in "parallel processing" fashion. Departmental installations can be individually configured and tuned according to the type of transaction being processed locally. Mainframe performance degradation is avoided as the demand for image storage and throughput increases. The aggregate physical imaging system hardware foundation is cheaper to scale-up with LAN-based implementations, and avoids these kinds of bottlenecks.



## 9.6 ISSUES IN IMPLEMENTING IMAGE ENABLING

### Workflow Control

A transaction's complexity determines the corresponding complexity of the workflow, but the mapping does not necessarily have to mimic what exists currently in the paper environment.

Many organizations initially configure the imaging system one way and then implement workflow redesign changes later once the system is in place and the operational staff has been fully trained. With intelligent workstation capabilities and access to multiple applications, an operator can be retrained to do in one step what may have taken several steps in the past. The intelligent workstation is a powerful operator tool that now can consolidate transaction solution data from widely dispersed sources.

### *Workflow Control—Mainframe or LAN*

With images residing on LAN-connected servers, data transmitted from the mainframe transaction processing systems can be used to control the placement and status of locally resident document images. The synchronization mode of mainframe and LAN processes depends in part on whether the transaction processing system is batch-oriented or on line, and on how complex the transaction is. Synchronization is more complex where many steps are involved in completing the transaction, especially if each step requires an operator possessing a different skill set.

### *Indexing and Referencing Document Images*

One of the critical decisions in image enabling a mainframe through a network is the location of the index to images. The problem is less technical as it involves strategic data management appraisal.

Indexing on the mainframe has the advantage of continuing the integrity of controls that characterize the mainframe. Its drawback is that it forces additional network traffic; in addition, as users migrate to multimedia applications at their workstations, centralized indexing may increase maintenance requirements to ensure compatibility. These issues may be difficult to handle with existing staff.

Local indexing offers more flexibility and is likely to reduce network traffic. At the same time, it raises the issues of cross-referencing document identifiers to the mainframe applications and maintenance of indexing standards.

### ***System Administration and Management Issues***

Integrating imaging with a mainframe also requires a review, from the imaging system point of view, of a host of items that probably were already resolved in the mainframe system. These include:

- Integration of remote diagnostics and alerts for all imaging operation problems, such as device failure. While some of these alerts may not be of interest to the mainframe, it is essential that their definition be compatible with those already in place. For instance, while it may be unlikely that a failed scanner that runs on its own network and under its own server control would be of interest to the mainframe, at some point in the future, fax machines and scanners may be the same input device, and the fax machine may be directed by the mainframe. Therefore, it is advisable that error messages be harmonized now to facilitate downstream maintenance.
- Configuration management and control, to ensure accurate cross-referencing between documents in image form, index, and their ASCII representation if any. This is especially important when documents with the same ASCII text are represented with images with handwritten annotations that reflect different versions of the same document.
- Security, from the point of view of access to both the installation and to the information contained in the system.
- Hardware and software management, to ensure compatibility in upgrades and maintenance.
- Archive management, to keep on-line storage requirements to a minimum and ensure proper disposition of obsolete records.
- Software distribution, to ensure—especially in a heterogeneous environment—that all upgrades are performed in a uniform manner.

In addition, close attention should be given to training and the coordination of staff development programs related to imaging systems, including users, operators, and systems administrators.

### ***Implications of Future Technologies***

The continued progress in hardware technologies, supported by increasingly accepted standards—including faster processors, fiber optics, width, multiprocessing, parallel processing servers, 32-bit operating systems, 64-bit processors, RAID, and Windows accelerators—will facilitate the integration of high-performance imaging systems with mainframes, but, at the same time, create additional pressure for downsizing.

These developments will persist in raising the issues of compatibility, cost, reliability, redundancy, and data integrity. Of primary importance will be the preservation of the various audit trails in the context of a changing technological environment.



## Checklist of Technical Issues

To facilitate integration of imaging with the existing mainframe, it is often useful to develop a checklist to consolidate the technical requirements of the mainframe operation and to query the proposed technical solutions to these issues. An example of such a checklist, which covers some of the "obvious and always forgotten" issues in system integration is given in Table 9.3.

## Standardized Integration Tools: Imaging and Mainframe Applications Working Together

Modern mainframe architecture provides tools to integrate imaging with existing applications, thereby eliminating the concerns about security and an audit trail that early file transfer protocols created.

Many of these tools are platform-dependent on the mainframe side. The use of off-the-shelf packages for the network side facilitates the integration.

### LU 6.2 and APPC

LU (Logical Unit) 6.2 is the underlying System Network Architecture (SNA) protocol suite that allows program-to-program communications across hardware and operating system platforms. APPC (Advanced Program-to-Program Communication) technically is the set of APIs that allows developers to code "conversations" between programs using the SNA network infrastructure. In this way programmers can build applications that cross platform boundaries without having to worry about lower-level communications details that are platform-dependent. A workstation can talk to a server process or a mainframe application using the same conversational verbs and rules. Mainframe applications, in turn, can activate server and workstation programs in an as-needed connect and disconnect basis.

APPC programming can be used to transmit local server processing statistics to a remote mainframe MIS reporting application, for example. As document images move electronically through different worksteps on the LAN, statistics for the rate of processing (to monitor productivity bottlenecks) can be easily collected and summarized.

Server processes can also converse with other servers, whether local or remote in the SNA network, using the same APPC syntax. The use of this standardized communication protocol for sophisticated control of the imaging system and related processes is the key to building powerful LAN-based imaging systems. It allows for the complex coupling of diverse applications on different platforms in many locations.

### APPN

Complementing APPC, which is already a well-established product, is the recently announced APPN (Advanced Program-to-Program Network) support, which now

**Table 9.3** Example of checklist of technical issues

<b>1. Intersystem dependencies</b>
Integration with multiple mainframe applications
Interdepartmental routing of work
Consistency of processing cycles with business needs
<b>2. Provision for growth</b>
<b>3. Provisions for scanning and archival operations</b>
Turnaround requirements
Capacity planning to avoid backlogs, day and night shift operations
Processing "windows" and system dependencies
High-volume print requirements
<b>4. Backup and recovery from component failure</b>
Procedures
Component redundancy
Software
Hardware
Network
Operational errors
<b>5. Data integrity procedures and controls</b>
Procedures
Audit trail
Data security and privacy
<b>6. Workflow management</b>
Procedures dealing with bottlenecks
Escalation of system problems
Peak versus average workloads system design
Workflow control
Manual intervention
Automated control
<b>7. File sharing</b>
Remote connectivity
<b>8. Education</b>
Senior management
Users
Operators
<b>9. Miscellaneous</b>
Facilities preparation



makes it easier for an enterprise to program, implement, and support LU 6.2 functions into applications without as much concern for network addressing details.

### Front-End Tools and Workstation Issues

As stated, imaging systems must deliver at least the same productivity as existing systems, subjective and qualitative savings notwithstanding. The workstation configuration is a key technical design issue; hardware power overkill can be very wasteful and costly, stretching the expected payback period out to unacceptable time frames. Completing technical homework in advance and shopping around can really pay off here. Recurring issues in the planning for system design and implementation include consideration of performance when replacing 3270 terminals with workstations capable of running up to five 3270-emulation windows.

#### *Windowed Workstations versus Dumb 3270 Terminals*

Operator acceptance of the new environment usually is very favorable. A "when am I going to have an imaging workstation" attitude develops, especially when the new system installation is accompanied by upgraded facilities and ergonomics improvements.

A windowed environment, because it is user-configurable, has the advantage over the 3270s of visual ergonomics. The availability of multitasking environments gives access to multiple applications and multiple transactions simultaneously, especially when assisted with front-end tools such as EASEL.

There are tools that allow a direct link to the mainframe screen. For instance, the IBM product known as EHLLAPI (Emulator High-Level Language Application Programming Interface) allows a developer to build links between a PC-based program and 3270 emulation screens. However, this API requires a significant amount of training, and its output is difficult to support; if a host application changes, so must the EHLLAPI mapping code. Also, EHLLAPI programming techniques are usually dependent on the anticipation of mainframe protocol behaviors (i.e., acknowledgments in the form of "not busy, you can send or receive data now"). Other more user-friendly tool sets are coming on the market to speed prototyping and development time.

In fact, protocols such as Windows DDE make it possible, with the proper workstation software and programming, to preserve a user interface while changing the attached mainframe applications, to the point where it is possible to substitute host systems without requiring any operator retraining.

Be aware, however, that there may be only a minimum productivity increase when developing front-end tools using interprocess communication protocols. A case in point is DDE (dynamic data exchange), defined for use in Windows and OS/2 environments, which, in complex applications, can be costly to develop and support without resulting in any significant productivity increase.

### **Image Retrieval Considerations**

The mainframe/LAN imaging configuration provides a simple model for document storage hierarchy. The work-in-process document images are stored on a local server magnetic disk to optimize workstation throughput until departmental transactions have been completed. The lifetime of the magnetic copy is usually very short; a corporation's daily mail receipts are typically entered within twenty-four hours.

For long-term archival needs, where legal requirements may dictate that the document image be retained for a long period of time (two to ten years), optical storage is the answer. Although microfilm is still cheap and suitable for many archival needs, as optical storage becomes more inexpensive and offers continual increases in storage density, the benefits of electronic access from remote workstations—document sharing, rapid retrieval, high-quality images, prevention of document loss, improved and highly controlled security—make optical storage the preferred method.

This model makes it possible to design an image-enabled system and to specify type and location of media, size of workstation memory and workstation hard disk, nature and location and size of servers' hard disks, mass storage devices, and optical disk jukebox (with mounted, unmounted, and resident platters, and unmounted and nonresident platters), as well as other mass storage modalities such as RAID arrays and backup tapes.

### **Other Considerations**

Other considerations include the handling of remote site image retrieval, WAN integration with imaging capabilities, and high-volume printing. The solution, which is application-dependent, hinges on the strategy adopted for mainframe-image integration, which includes the choice to incorporate either local or mainframe indexing. The applications can be programmed to run either on the mainframe or from the local workstation, either using all necessary local and remote network resources or the mainframe as a node on that network or on the mainframe.

Image enabling requires a review of existing configurations and the applicability of monitoring and reporting tools, system administration, and integration with the existing network attached to the mainframe. In particular, there will be questions regarding:

- Methods of communicating (addressing) with departmental servers and workstations
- Benefits of local LAN versus WAN requirements or bridged LANs
- Geographical dispersement bandwidth considerations, and contention with imaging traffic in various topologies, such as the 16MB Token Ring
- LAN image location and storage management
- Loose versus tight applications coupling
- Mainframe batch processing control
- Impact of imaging in 3270 traffic and response on the LAN



## **LAN/WAN Integration for Dispersed Geographic Sites**

While there are numerous tools now available to perform integration with existing mainframe-based systems, the real potential lies in enhancing these systems to provide a tighter coupling with the LAN-resident imaging systems. Workload management based on transaction processing rates and attributes and sophisticated linear programming techniques may be implemented in the centralized mainframe application in order to more efficiently control the placement of images on LAN servers in anticipation of their use by a certain department or even a particular workstation. This requires a combination of mainframe integration and client/server-based programming techniques. It can provide maximum productivity by caching images locally in anticipation of their need. In this way, image retrieval response from local hard disks takes seconds, instead of jukebox/optical platter retrievals, which can take hours depending on the status of the platter and the volume of pending requests.

## **Backup and Recovery**

Backing up an image-enabled system is not a trivial issue. Beyond the physical backup of the stored images, done by disk mirroring or off-time replication, there are issues of synchronization with the mainframe applications that must be resolved. While there are several tried and true mechanisms for ensuring rapid recovery in traditional mainframe applications, the integration of a similar level of data integrity for the LAN-based system may have to be developed internally.

In particular, the inability to recover quickly from individual server failures can cause major system problems.

Recovery time must be as quick as possible to ensure critical and high-availability operations. The adoption of advanced LAN-based imaging software with "failsafe" processing, where a single departmental server failure will not affect any workstation processing, is one way of achieving this goal. Other solutions may lie in the choice of the topology of the network itself: Bridges and intelligent hubs make it possible to bypass failed devices; additional protection can be implemented through the use of RAID disk arrays and fault-tolerant server architectures.

Whereas a downsized application is independent of any mainframe processes, even though the data are being accessed, an imaging system integrated with a mainframe application brings a new set of problems. The two applications can be processing units of work independently or dependently. When there is transaction dependency between the applications, this implies that the two applications are synchronized. For example, the scanning of a new image and the creation of its associated index information on the LAN is usually followed by the transmission of this information as "initial inventory."

The systems must therefore maintain their synchronization for accurate processing of transactions, reporting on work in progress, and so forth. The two systems can easily get out of sync if one of them loses critical production data (i.e., a server crash or mainframe database loss in the middle of the day). Both applications may

possess backward and forward recovery facilities, but recovery may require an additional step before resuming production work: verification of synchronization.

### **Additional Consideration When Downsizing**

Downsizing efforts (or, more accurately, right-sizing) are often considered in parallel with the implementation of imaging systems in a mainframe environment. When this occurs, several conflicting considerations must be taken into account.

In traditional mainframe environments, the CPU performs processing for all terminals. Traffic to users is limited essentially to characters, fields, and print files, and the heavy data traffic involving transfer between storage and CPU is done on dedicated channels. In a downsizing environment, that traffic appears on the general network, claiming its share of bandwidth. Considering that a mainframe may communicate at 4.5MB per second—about 50 percent of the bandwidth of an Ethernet channel—this is sufficient to choke any attempt to run imaging applications on the same network.

The solution to acceptable service requires reducing traffic on a given network segment or increasing available bandwidth or both. Techniques include data compression, increasing the number of file servers, use of different indexing techniques, and network architecture redesign. As indicated earlier, the use of simulation of the proposed installation is essential.

---

## **9.7 CONCLUSION: OUTLINE OF IMPLEMENTATION STRATEGY**

---

The following list provides a model for the implementation of an image-enabled mainframe system:

1. Complete feasibility studies and functional specifications.
2. Identify and resolve interface issues between the mainframe and imaging network.
3. Complete technical specification.
4. Select imaging system vendor or in-house development team.
5. Prepare facilities.
6. Install network, server, and workstation infrastructure (hardware and software).
7. Train departmental users and imaging system support personnel.
8. Complete user acceptance testing and sign-off.
9. Convert from paper and microfilm to high-volume scanning, archiving, and retrieval operations.



10. Migrate to imaging systems by line of business, associated with corresponding mainframe application, using terminal emulation and image display at the workstation.
11. Upgrade to enhanced workstation capabilities.
12. Upgrade from loosely coupled to tightly coupled applications integration.

# Database Structure for Document Image Retrieval

---

*Jeff Berger*

Motorola

*Oren Sreebny*

BRS

## **Editor's Notes**

*Databases are at the heart of any information system, as organized data repositories. That organization is essential to efficient retrieval and management of documents, regardless of their medium and format. Supporting a document database are other databases hidden from the end user and maintained by the computer platform to manage the operation of that platform. The database that is of interest to us in this chapter is the only "visible" one, used for document storage and retrieval. To be usable, the database must be queried by a "database engine," which interprets user requests, locates relevant data, such as address of documents matching a query, and passes the results to a retrieval and display process.*

*There are different types of database engines, each corresponding to a design trade-off made to reach the optimal compromise between various operational factors, such as speed of access and use of computer resources to support a given application.*

*Consequently, while all "reasonable" database engines could probably be used in any application, some are better suited to one simply because*



they are more efficient in that given environment. For instance, a Standard Query Language (SQL) engine can well serve applications represented by sequential files; however, such engines would not be as efficient as a simple indexed database engine. Similarly, optimal information retrieval solutions for heterogeneous data formats and arrangements, as for imaging systems, will require more complex database engines.

Because imaging inserts large blocks of non-machine-readable data into what may be an otherwise fully searchable database, the structure of the database used for image and data retrieval must accommodate demands on both the logical (i.e., the contents) and the physical (i.e., the form) aspects of a document. In particular, the database structure of the retrieval system must at least be able to accommodate retrieval of document images and of related indexing data, and at the same time be able to support requests from users coming over the network, without impacting on the overall network performance. One approach, which is practically universally accepted by now, consists in isolating logical from physical retrieval. This is done by performing a query—not on the large computer files containing the documents, but on the smaller database, which contains the index to these documents—and passing to the retrieval process the address to the document rather than the document itself. Even when queries are done through full search, rather than through an index, the search is performed on a surrogate file containing the textual information, the size of which is smaller (and definitely more machine readable) than the image database. To ensure that heterogeneous systems can communicate over a network with minimum amount of programming, most modern database engines provide application programming interfaces (APIs), which are specifications for their interface with the outside world. In that way, a queried database must respond only in the appropriate API format, without concern for the mechanics of the data processing activities beyond its interface. At the same time, the application—such as the software driving a jukebox to retrieve a particular record—will only need to be programmed to understand the interface commands, and will not have to be concerned with algorithms used in the database to provide the proper identification of suitable records.

This chapter discusses the various elements that enter into the construction of an imaging system database and how they relate to a network windowed environment. In particular, it discusses some of the issues involved in integrating document retrieval with other applications in a multitasking environment. It will be shown that construction involves considerations that are common to the design of the overall system, such as legacies and organizational policies. Some of the material given here may repeat segments of information in other chapters, but this redundancy has been kept to ensure complete treatment of the subject matter.

*What needs to be kept in mind is that, regardless of the progress of technologies, database structure is a design issue rather than a technical issue. In particular, it must be understood that the inefficiencies of a given database design may be initially minimized, or even hidden, by the use of faster processors. However, each system has its inherent limits, and it is imperative that the system designer remain sensitive to their existence, so that they are not reached before the system has exceeded its life cycle.*

## 11.1 ELEMENTS OF THE IMAGING (OBJECT) DATABASE

### Introduction

The fact that large blocks of stored data carry with them no extrinsic information sets this kind of system apart from traditional data environments. To allow the cataloging and finding of these binary "dumb" objects, it is necessary either to identify external data elements or specifically create them. In other words, the difference between a text and image database is that the image (information) does not usually contain searchable data elements; external data elements must be added and linked to the image to provide contents that contain both the image and information about it and its location. This creates what is sometimes called metadata elements, with the image portion making up the single largest element of an image database.

Another difference is that traditional databases, while often made user friendly by a graphical user interface (GUI), do not necessarily require more than a simple ASCII terminal. This is not true in an image-based environment where not only may the user interface be graphical, but the data are graphical and demand an inherently graphical environment. Also, file systems are naturally larger in an image-based system; a page of ASCII text is about 2KB in size, and a Group IV compressed image of the same page takes approximately 30KB of storage. Whether optical storage systems are required over a more traditional magnetic medium will be discussed later. Finally, as will be discussed here and elsewhere in the book, the demands on the network environment are greater with images than in more traditional databases due to two factors: The nature of the data and the traffic associated with both host environments and client/servers place high demands on the network; second, and perhaps equally significant, is the desire of the system developers to expand the imaging system to a user base not normally associated with a traditional system, or at least not with the same system. The addition of purchasing, material quality, manufacturing, sales, engineering, and even management expand the scope, and, hence, the network demands of an image-based system.

### Database Models for Imaging Applications

The dominant conceptual model in current database technology is the relational model. However, to accommodate the differences in information contents between



machine-readable data and images, it is often necessary to modify that model. The following compares a traditional relational model and one that has been enhanced for images.

### **Relational Database Model Overview**

Relational database systems can be described in terms of two rules:

1. The data are perceived by the user as tables (and nothing but tables).
2. The operators at the user's disposal (e.g., for data retrieval) are operators that generate new tables from old.

One classic example of a relational database implementation is for tracking inventory and sales in a retail store. Items for sale are usually distinguished by an identifying item number, while customers are referred to by a customer number. We could construct tables for inventory, customers, and items as shown in Tables 11.1–11.3.

It is easy to construct a query to retrieve a list of names, addresses, and items for delivery of sales. The query could look something like:

```
SELECT
    Customer.CustomerName
    Customer.CustomerAddress
    Item.ItemName
FORALL Sales
```

The table resulting from our query would look like that shown in Table 11.4.

This very simple example serves to illustrate some of the attractive features of a relational database system. It offers a tremendous amount of flexibility to the user of the system. Tables can be combined at will, based on key elements (such as Item# and Customer# in our example). In addition, the complete descriptive data for each data element, such as the customer's name and address, need to be stored in only one place in the system. In addition to saving storage, this makes maintenance of the system considerably easier.

It is important to remember that relational databases are optimized for flexibility, and, as such, have potentially significant shortcomings. The relational model typically requires a high degree of structure in the data it stores. While this works well for many commercial applications, it is not so advantageous when structuring the data in a document-based store. Retrieving elements from a relational database can be time-consuming because, typically, the items within a relational table are not ordered. To retrieve a specified item then, it is necessary to step through the entire table one row at a time, comparing each value to the specified search attribute. It is possible to set up indexes for columns in relational tables, but this is not commonly done for all data elements.

Table 11.1 Item relational database

<i>Item#</i>	<i>ItemName</i>	<i>StockQuantity</i>
001	Dining room chair	5
002	Dining room table	2
003	Double bed	4
004	Baby crib	1

Table 11.2 Customer relational database

<i>Customer#</i>	<i>CustomerName</i>	<i>CustomerAddress</i>
001	John Martin	80 Westpark Dr
002	Peter Philip	1301 Circle Rd
003	Jack Smith	455 Manor Rd

Table 11.3 Inventory relational database

<i>Sale#</i>	<i>Customer#</i>	<i>Item#</i>	<i>SaleQuantity</i>
001	002	004	1
002	001	001	3

Table 11.4 Query result

<i>CustomerName</i>	<i>CustomerAddress</i>	<i>ItemName</i>
Peter Philip	1301 Circle Rd	Baby crib
John Martin	80 Westpark Dr	Dining room chair

### Document Retrieval Database Properties

In a relational database system, retrieval is a one-step process, generating a new table that represents the results that satisfy the query. In a document retrieval database this is a two-step process: In a search operation, the user enters queries into the system, which then identifies a set of documents that satisfy the query; the retrieve operation is then issued against the identified set of documents, which are then presented to the user.

The types of documents that may be stored in a document retrieval database vary widely. Document retrieval applications might include storing technical documentation, procedures manuals, memoranda and correspondence, legal depositions and briefs, published journals, and library catalogs. Some of these document types may



be highly structured. Memoranda, for instance, typically contain an opening section where the addressee, sender, date, subject, and distribution list are specified. Similarly, manuals usually contain hierarchically organized chapter, subchapter, and paragraph headings. What is common in these and most documents is that the actual body of the text is made up of unstructured prose (although we should not assume that document retrieval databases are limited to prose; see, for example, Dartmouth University's Dante and Shakespeare databases). This text is not only of variable structure, but is usually variable in length as well. The storage and retrieval of these documents is the goal of the document retrieval database.

What are the properties of a database system for retrieving this wide variety of documents? As with the previous definition of a relational database, we can specify two rules that define a document retrieval database:

1. The data are perceived by the user as documents.
2. The operators at the user's disposal (e.g., for data retrieval) are operators that identify and retrieve a qualified subset of the entire document store.

This definition implies a basic feature of document retrieval databases—that the elemental unit of retrieval is a single document. This means that the results obtained from a query of the document database are entire documents, even though in a particular situation we may choose to display only certain portions of the document to the user.

The definition of what constitutes a single document may differ from one database to another, even if the databases contain similar material. One database containing published journals, for instance, may be structured with a single journal article as a single document, while another database may be designed with an entire issue of a journal as a single document.

### **Database Structure for Document Retrieval**

As already noted, the elemental unit of retrieval within a document database is the individual document itself. But what are the logical file structures used to obtain this retrieval? While the standard (relational) database is composed of records and tables, the image database is composed of an additional element: the image object. Even though an image object can be found in a form such as a Computer Graphics Metafile (CGM), which allows direct inquiry of the object, most imaging systems must be designed to allow for pure bit-mapped or raster data. To the database, these are completely "dumb" objects.

To represent that variety of information objects, the most common data model for document retrieval databases is the inverted index structure. The inverted index model, developed in the 1960s for the maintenance of large indexes to scientific and commercial literature, is used in most of the large commercial information retrieval services, such as BRS, Dialog, Mead Data Central, and the National Library of Medicine's Medlars service. And while they are widely available in many commercial document retrieval software systems, inverted index database

models, which are highly effective in the retrieval of unstructured information, are not optimized for flexibility. Indeed, the structure of these systems frequently requires complete reconstruction of the indexes with each database update. The inverted file structure, however, does offer extremely rapid database searching that is not very sensitive to the size of the database. This allows the construction of powerful mixed-format databases that contain millions of documents in both image and machine-readable format.

### **Archival, On-Line Reference, and Transactional Applications**

Database structure is strongly influenced by the purpose of the information system. Most systems are optimized for a particular type of application, such as archival, on-line reference, or transactional. Because imaging applications will span the life cycle of business records, enterprise-wide databases may need to be segmented into local databases, optimized to provide the best response to a given set of operational circumstances.

### **Indexing Methodology**

Another decision to be made in the selection of a database structure is how to index information for document retrieval, a subject that has kept librarians and classification specialists in business for hundreds of years. The advent of computerized information systems has provided very powerful tools for enhancing the capabilities of document retrieval systems. It is still true in retrieval system design, however, as in most computer systems work, that a little thought and effort in the design stage can save untold amounts of work and hardship in the implementation.

### **Language and Indexing**

In a document imaging system the actual retrieved document will be, by definition, an image. This may be an image of a textual document or it may be a graphical image. Whatever the actual image represents, in almost all cases the method used to describe this image for retrieval will be the written language. We do not, for example, describe a photograph by drawing a picture of the photograph, but by describing its content, color, and/or shape ("a large red balloon"). It is the job of a document imaging system to match the user's linguistic request against an index of linguistic descriptions of documents to produce a document image.

**Imprecision of Language** The use of language is notoriously imprecise, and terminology shifts among nationalities (the English auto's "boot" is the American car's "trunk"), over time (the "trash collector" becomes the "sanitation engineer"), and between different professions (the word "program" has very different meanings to computer professionals and television executives). In addition, the proliferation of nonunique acronyms (is the NEA the National Education Association or



the National Endowment for the Arts or both?) and shifting usage make linguistic information retrieval a mine field of problems for the system designer. Even seemingly simple data elements such as personal names can pose problems: Are memos addressed to Bob Briggs and Joseph Robert Briggs to the same person? In addition to the natural imprecision of varied living uses of language, well-written documents are not necessarily well suited for retrieval applications. When we write prose we are taught to vary the ways in which we refer to things; we are instructed not to express concepts in the same terms over and over again. In document retrieval, however, we would prefer to have the same terminology used consistently in every instance so that we may retrieve all documents that refer to a given concept with assurance.

**Tools for Dealing with Linguistic Problems** There are tools available to help the database designer cope with this linguistic morass. The most rigorous method is to design and enforce the use of a thesaurus to control the indexing vocabulary. In a thesaurus the database designer decides upon the approved term for each concept to be represented in the index. These approved terms (known as the "preferred terms") are the only representations for these concepts that will be allowed to appear in the index. The thesaurus designer then lists all other terms used to represent each concept and builds pointers from these "entry" terms to the preferred term. The use of this thesaurus file can be automated to normalize data upon input as well as to map search terms at retrieval time. Other tools for coping with linguistic variations make use of word stemming, automatic pluralization, and phonetic algorithms.

### ***Full-Text Searching***

Most current document retrieval systems support the indexing of all of the text in a document. These are referred to as "full-text" retrieval systems. The power realized from having each word in a piece of text indexed is immediately apparent. In order to support full-text searching, document retrieval systems allow the user to specify sophisticated contextual relationships between multiple query items, such as requiring that terms co-occur within the same sentence or paragraph of text. The database designer must decide to what extent the complete text of documents will be indexed. There are costs involved in making the full text searchable, including the cost of converting document text into searchable form and the additional cost of storing the extra indexing. In addition, the written text of most documents is exactly the type of material most prone to the linguistic retrieval problems discussed above. Full-text indexing usually implies that the material being indexed was not written and composed for the purpose of enhancing information retrieval. Because of this the user may end up retrieving documents that are only tangentially relevant to his or her query (because the query terms are

concepts mentioned only in passing in the document) or miss relevant documents altogether (because the document author used different terminology).

### ***Controlled Indexing***

The system designer can alleviate the problems inherent in full-text indexing by either utilizing the existing document structure or by adding structured elements to documents. If, for example, the database is to consist of business memoranda, the database designer can place the DATE, FROM, TO, and SUBJECT data in separate search fields from the body of the text. This then allows for greater specificity when constructing queries. In addition, the system architect can choose to design query structures that weigh terms that occur in certain fields, such as the SUBJECT field. The database designer may decide that a given document base does not possess enough structure to make efficient retrieval possible. In this case, the addition of descriptive indexing to the original data may be warranted. The terminology used in such descriptive indexing should be controlled as much as possible to ensure consistent retrieval. The designer may be able to use existing lists of terminology from database publishers, such as the National Library of Medicine's Medical Subject Headings or the Institute of Electrical and Electronics Engineers' INSPEC index.

### ***Indexing for Image Retrieval***

While it is common for an image to represent text, it is important to remember the difference between actual text (as represented by ASCII character codes) and the image of text (this distinction is analogous to Magritte's famous surrealist painting of a pipe with the legend, "This is not a pipe"). This distinction may seem obvious to computer professionals, but it is one of the most common points of real confusion in system users. The advent of GUIs and compound document architectures may serve to further cloud this distinction in the minds of system users. Since the image is not made up of text, it follows that the actual data being searched are not the image. In other words, while the data that are indexed may have been derived originally from the document image, the image itself is not searchable. In most imaging systems the ultimate view of the document presented to the user is an image of a document instead of an ASCII text version of the document. This means that the indexing must lead the system to the relevant image when a display is requested by the user. The image data may be stored within the actual document retrieval database itself, or they may reside external to the database in separate image files. If the image data are stored in separate files, the document retrieval database must store pointers to the locations of these files. Since the system administrator will want to remain flexible in choosing file locations, the system designer should be able to place an additional level into the database design, so that absolute file names and locations will not need to be stored in the actual database.



---

## 11.2 DESIGN REQUIREMENTS OF THE IMAGING DATABASE

---

Experience has shown that a well-designed document imaging database requires consideration of the four main system elements: document capture, control/indexing, file system, and document display. Other elements can be added, such as OCR, and fundamental system issues such as backup and account management cannot be ignored, but those four elements are the cornerstones of any imaging system.

When laid against an emerging set of requirements, these four elements can help to crystallize a focus for the design of the system database. When broken down from all the bells and whistles, there are some very simple and fundamental items to be considered.

### Overall Considerations

#### *Document Capture*

What is the source of the documents: paper (via scanning) or electronic (direct conversion of an existing electronic form)? Will an existing archive be converted, or does the system proceed from the day of initiation? Will the system scan or convert at a central location or multiple centers, or will all users be able to enter documents? What are the control points and checking associated with entering documents, and what are the criteria for saying they are accurately converted and indexed? Who signs off in that approval process? If a process view of the entire imaging system is used as a check against proposed requirements, flaws in the thought process are revealed, and often, more importantly, overdesign of the system can be avoided. We will come back to overdesign later.

#### *Control/Indexing*

What is the minimum indexing needed? What cost in time, hardware, and software is required to achieve that level? How much training is the organization willing to pay for both users and administrators? Can the indexing scheme be modified easily and at what cost? What is the expendability of that scheme? Does a new workflow need to be established, or can the system achieve payback goals by merely getting out the paper?

#### *File System*

How much information is reasonably going to be generated? How long do data need to be kept on line? What amount of backfile conversion must be accomplished? What response time is required for document access, both recent and

historical? Can the data be maintained on magnetic media without optical storage? What network access must be maintained? Does the backup strategy meet with corporate and project requirements for frequency and off-site storage, particularly in the case of optical-based systems? This category is a great place to meet head-on with the issue of overdesign and out-of-control costs.

### **Document Display**

The selection of a display system or viewer is key to user acceptance of the imaging system; it must match the function performed. The world's best-in-class viewer for check-clearing applications may be totally inappropriate for engineering drawings. Questions to pose include: How many hours a day will the user be in front of the system? Will the average user be performing other functions on the computer? What is the time to invoke versus the average length of session? Are hardware accelerators needed to meet productivity goals? Is this system used in a workflow or documentation application where annotation is required? As above, what cost in training is the organization willing to expend per user? What is the user turnover and the related impact on training? What level of expertise and time is required to administer the end-user software? How is it licensed and maintained, and what impact does that have on long-term cost of ownership? The temptation to allow all possible combinations and permutations of options just does not stand up to the real world in the way of the source and number of documents and the skill level and training of the people using and maintaining the system. The 80/20 rule must be applied in avoiding system overdesign. A system that meets 100 percent of the initial system requirements 100 percent of the time probably is not affordable or practical.

### **Other Considerations: Cultural Elements of Systems Design**

Because the imaging system will so dramatically change the business, the system, and, consequently, its database, it cannot be designed without external consideration of the information process itself. In particular, the design and installation of an image management system and its underlying database involve not only technical but also cultural decisions. Integrating system analysis, process reengineering, organization mapping, and hardware and software selection is a significant technical challenge. However, the success of the imaging system in solving a business problem and achieving productivity improvement is only partially impacted by those technology-oriented decisions. To repeat the prime directive of imaging systems: Putting images into a document management system has no payback whatever; only retrieving images from such a system results in productivity. The point in this case: Users must perceive an improvement in their job function before they will use it. Until someone uses a system it will not be productive and will provide no payback. Thus, managing the user's perception of the document



management system is probably the major element in successful implementation. Successfully managing the user's perception of the system will make the most cumbersome, archaic system a success in his or her eyes and, hence, in that of management. Simply turning users loose without sufficient support will sink the most elegant solution absolutely.

Books have been written on managing organizational change,<sup>1</sup> and an extensive treatment of the subject is outside the scope of this book. Nevertheless, some key elements of managing the cultural aspects of imaging system installation as they directly or indirectly affect the success of an imaging system must be discussed.

### ***Involvement***

Users must be involved in the entire process of developing an imaging system, from selection of projects to concept definition to final sign-off. If, as in many cases of imaging system design, a committee is making decisions, users (end users, not administrators or support staff) must be the key contributors to the process. If they do not understand the problems, then they must be educated first.

### ***Ownership***

Very similar to involvement, but beyond mere participation, users must be made to feel responsible for the outcome of the project. Involved users who don't feel a sense of pride and responsibility in the effort will have little motivation to reach deep and explore the possibilities of the new system. Lack of ownership will result in responses to the pilot phase of a project such as, "I know I told you that's how we do our work, but what really happens is. . . ." This kind of organizational behavior can turn a four-month project into an eighteen-month project before you can say, "Where's the schedule?" An almost infinite number of three-day and seven-day delays crop up out of nowhere. Anyone currently managing an out-of-control project who can't really pinpoint a specific problem would be well advised to look at the organization and its involvement and ownership, not of the developers, but of the end users. The symptoms are: a constant stream of changes, requirements well beyond any that your system or imaging supplier has ever run into, constant refocusing of the project, and routine turnover of "dedicated" resources from the end-user community.

### ***User Perception of Management Commitment***

The subtle difference between managing the user's system and managing the user's perception of the system cannot be overemphasized. Peters and Waterman in *In Search of Excellence: Lessons from America's Best-Run Companies*<sup>2</sup> clarify that the difference between the delivered quality and the user's perception of quality is the final differentiation in reaching the highest attainable system quality. At first, this may sound like a way to deliver a less than high-quality system and "sell it" as quality, but it is not. No amount of selling will convince an end user that he or

she is receiving a quality product if it does not perform. After delivering a quality system, it is time to begin managing the user's perceptions: Is it well supported? Is it available? Was I trained adequately? Can I see the payback? Is my life better? Are my customers better served? When the user answers these kinds of questions in the affirmative, a good system has been delivered. In the list of key elements above, the user's perception of management support hits directly on this issue. If users perceive that management believes the project is critical to success, they will be encouraged to contribute to that success. If management is committed but the users don't know it, either through lack of communication or visibly inadequate support, the project will suffer. If management truly does not support the project, why is it happening?

---

### 11.3 PREPARING THE IMAGING DATABASE— PRELIMINARY QUESTIONS

---

Once the overall structure of the database has been selected, the next priority is to refine the imaging system design by supplying answers to the following questions.

#### **What Information Is to Be Stored as Images?**

Choosing to store information as an image can be a complex decision. If information is mostly text with few embedded graphics, cost-effectiveness may drive a decision to file away the graphics and rely on the text alone. The wisdom of this depends on the impact of not having easy access to the graphics, the time and distances involved in recovering a copy of the original paper, the desire to get rid of paper completely, and so forth. In essence, ask what the difference is between the cost of making a wrong decision by having access only to the text of a document and the cost of storing sparse graphics in an otherwise text-only database. If access to source electronic documents is possible, conversion to a vector/ASCII form such as CGM may be ideal. A high content base of paper archival, inbound paper, and CAD-based documents (where non-CAD users need access) indicates image forms should be incorporated.

#### ***Textual Information***

There is little reason why text should ever be stored as image (raster) since most image viewers can also view text. However, the capture of text data requires some thought as does the format of the text itself. Unformatted text, formatted text, and various forms of propriety text such as Interleaf ASCII, Encapsulated PostScript, and CGM are all possible ways that text information may arrive. Retention of descriptive data (font types and sizes, tab widths, underlines, etc.), conversion before storage to a common form, or conversion upon retrieval are all options that will place points on the cost/performance curve. Should you choose to retain



descriptive information, the impact on a viewer that handles all of the above may be more costly from a purchase price and training standpoint than converting everything to a formatted ASCII for storage and viewing.

### **Graphics**

The kind of graphics that will be stored is another of those areas where the 80/20 rule needs to be carefully evaluated. The decision that everything from CAD drawings to PowerPoint to full-color photographs must be stored can be the most expensive one any system buyer can make. The cost in conversion utilities, hardware, disk storage, and viewers can send an image system budget ballistic. A key to successful implementation is in narrowing the focus to one that satisfies a current, well-defined need and is expandable to meet future real or imagined requirements. This point cannot be overstressed: Meeting a current need is the key to initial user and management buy-in to an imaging project. Systems can be determined to be expandable in several ways:

1. Existing capability: The system being purchased already can perform the functions required in planned expansion phases.
2. Future capability: The system vendor can credibly show that the system purchased will have the future functionality to meet expected growth.
3. Conformance to standards: The system installed conforms to sufficient standards for image format, storage, and indexing methodology so that merging new functionality would require minimum energy.

### **How Is the Information Currently Stored?**

There are several areas from which the image and indexing information can be extracted. System design must zero in on the top hitters for efficient movement between environments.

### **Paper**

The elimination of paper accounts for only 25 percent of imaging system installations according to a recent Delphi Consulting Group survey as opposed to cost cutting and competitive advantage. Still, if that is the user's goal, decisions about back-converting (capture of existing documents) and handling of new work must be made (see above), and the process of rooting out and eliminating paper must be addressed. Management of paper-based documents during conversion must be carefully handled as well, as changes occurring to documents during scanning can cause obsolete information to populate the database. The document management system, either a new one or a modified existing system, must be capable of managing the transition period while both paper and electronic media exist. In some cases where the cultural barriers are particularly high, it may be advantageous to

run both systems in parallel and let users migrate gradually to the new system. Initially, the "techies" will switch, taking the role of early adopters, and they will provide encouragement and training for the less willing. Gradually, management must increase pressure to change over until the final day of cord cutting. This dual system is easier to maintain in a document distribution environment such as an engineering printroom more so than in a transactional system such as insurance claim processing.

### **Electronic**

When the source of a document is an existing electronic form, other issues need to be addressed in formulating an effective strategy for image management. The foremost is perhaps, "Does this need imaging in the first place?" Department-level organizations or centrally managed enterprises must discover whether documents can be managed and shared using a common software platform, such as converting all to Word, Excel, or MacDraw. Since most departmental users require some or all of those tools anyway, document sharing and management can be done with simple file management functions from a product such as SoftSolution, rather than a complex electronic image management system and associated conversion tools.

**Word Processing** The conversion of word processing documents into an image format can be easily accomplished. There are converters for PostScript output into a variety of raster and CGM formats. There is little case for the conversion of unformatted ASCII text to an image format. An analysis of where the documents will be used, how they are transmitted, and the volume will help determine whether the documents should be stored in the native WP format, include computer graphics metafile (CGM), or down to raster images. The mix of this kind of document versus others such as large-format engineering will determine the image viewer required to complete the system.

**Spreadsheets** Spreadsheet data are a special case of the above in that most users will have a spreadsheet program of their own available, obviating the need for imaging. In the instance where the data are to be distributed but not in modifiable form, conversion to image may be desirable, but, as in the following case, it may require a viewer that supports documents larger than the standard legal size.

**Project Schedules** The distribution of project schedules via image to large-format viewers with annotation capability can be a key way to implement a two-way, documented discussion of project management issues. In this way, a "frozen" version of a schedule can be released to a large class of project personnel, other impacted individuals, and management for review and markup without requiring that everyone purchase and be trained on the scheduling tool. The conversion to image means that users can be contacted via e-mail or other electronic means, which makes it possible for distant locations, such as vendors and remote manufacturing sites in one application or co-counsel attorneys in other applications, to



be kept "in the loop" on project changes when paper would be slow at best. Large projects rarely fit on an 8.5" x 11" page, so that form of contact is even more cumbersome. Conversion through capture of a print spool file can be an easy way to configure and convert such large-format documents.

**Databases** The use of imaging to distribute database reports would seem to be of little use when, as mentioned above, most image viewers can display text and some even support the graphical annotation of text files, such as Motorola's WaveSoft.

**Graphics** The distribution of graphics via imaging systems, whether stand-alone or embedded in another document, is what we are all here for. The selection of viewers, conversion methodologies, and compression and decompression techniques all must center around the type of graphical objects and the end user of such documents.

### **How Much Information Is to Be Stored?**

When the sources of data are so numerous and those sources have the ability to generate unprecedented quantities of data, the system designer must carefully evaluate what to save and what to let go. The primary issue is cost: the cost to capture or convert images, the cost of magnetic or optical file storage, and the cost of ownership such as backups and maintenance. Other factors are also significant in this decision, but honesty with oneself and the users is probably most important. It's important to ascertain what the chances are that anyone will ever use the information inscribed versus the cost of storing or recreating the information through other means.

We have seen manufacturing environments that were entirely paper-based convert to imaging. The simple choice would have been to save all versions of the drawings on line. But examining the current system revealed that when paper was the master document, the original vanished when the pencil lines were erased and the new version was created; thus, the current version was the only one easily accessible, and everyone was used to that. No quality or end-customer factors could be found to justify making the on-line system larger to accommodate old versions. Therefore, the urge to save everything on line was repressed, and the system functions very well with off-line storage of old versions. Of further impact on the decision to store only the current version was that magnetic storage was sufficient to hold the information, thus cutting substantially the cost of the file system over the optical system proposed. Additional ripple effects in the maintenance cost, ease of backup, and access time made a further positive impact on system performance and user acceptance. As discussed in the next section, it is possible to put more information in an image system than is practical to index unambiguously. The indexing capability selected should be evaluated when deciding how much to put into an imaging system. System developers must be honest about how items will be found, as well as if anybody really wants it.

## What Kind of Retrieval Is Needed?

The design of a retrieval system should be based upon the actual user requirements of the particular situation. The system designer must take a careful look at what types of retrieval are necessary for a given document store in order to design the database to perform adequately at an acceptable cost.

### *Recall and Precision*

The effectiveness of textual retrieval is often measured in terms of recall and precision. Recall measures the proportion of relevant documents retrieved by means of a query to the relevant documents not retrieved by the same query. Because of the difficulty in calculating the absolute number of relevant documents within a large production database, recall measures are usually performed on controlled test collections with controlled queries. Precision is the measure of *relevant* documents retrieved out of *all* of the documents retrieved. Because precision is calculated on just the retrieved set of documents, it is not necessary to limit its measurement to a test collection. Precision and recall are often portrayed as being inversely proportional: If measures are taken to increase recall, decreased precision will result. While this is often true, there are measures that can be taken that increase both precision and recall, such as improvement in the consistency of indexing or by overall improvement in the structure of a query. The absolute measures of precision and recall are entirely dependent on the contents of both the database and the query. These absolute measures are seldom important outside of laboratory settings. In a given design situation, however, it can be helpful to conceive of the system as emphasizing high recall (for instance, in patent searching applications where missing a single document can be crucial) or high precision (as in a library setting where students are likely to be looking for a "few good references").

### *Keyed Retrieval*

In some situations, a unique document identifier may be the only search key necessary for image retrieval. This may apply, for example, in parts inventory systems where the only search criterion needed is the part number. The attractions to limiting to this type of retrieval are its simplicity and low cost. The disadvantages are in its relative lack of flexibility and power.

### *Descriptive Information*

The next step in retrieval complexity and power is to add descriptive information to the document retrieval index. As mentioned before, the more the descriptive information can be normalized in a consistent manner, the greater the usefulness. Even in the simple example of the parts inventory system mentioned above, a



phrase with the part name and one describing the part can greatly enhance the power of the system.

### **Keyword Retrieval**

Keyword retrieval is the indexing of each separate word of a body of text. This body of text can be simply descriptive information or the entire text of a document. The power of keyword searching is harnessed by indexing the location information of the keyword, allowing the use of positional operations such as word adjacency or nearness. The ability to specify terms occurring in the same sentence or paragraph is another common positional operator.

### **Full-Text Searching**

Full-text searching is the most complex and powerful method of document indexing. In full-text searching every word of a document is indexed for searching, with the usual exception of some words that occur so frequently as to render their informational content null (such as "the" or "and"). As mentioned above, full-text indexing is the most costly and the most complicated type of indexing to prepare. There are, nevertheless, situations where no other type of indexing will do, or where the effort of full-text indexing is easier than going through the rigors of preparing methods of controlled document descriptions.

---

## **11.4 PREPARING THE IMAGING DATABASE— OPERATIONAL QUESTIONS**

---

The imaging database design is also influenced by the conditions under which the system is going to operate and the record life cycle. The following questions need to be answered before a design can be finalized.

### **How Will the Images Be Captured?**

In developing the environment for document capture, the two key questions to investigate and decide upon are: "what" and "who." The "what" is answered by paper or film, electronic, and both. Existing bit-mapped documents such as paper and microfilm have few choices for conversion and those revolve around the choice of scanners and whether to develop an in-house scanning capability or go to a service bureau. Scanners fall into two classes by price: those that require some amount of adjustment to get a good image and those that are compensated for more or less automatically. The newcomer is faced with this dilemma: A desktop scanner in the \$500–\$1,500 range can scan at up to 600 dpi (dots per inch) and has a sheet feeder. So why pay \$10,000 and up for a scanner with the same capabilities? The user will discover that unless all the documents come off a laser printer and go

directly to the scanner, the inexpensive system will require multiple scans of each page to get the brightness and contrast settings to a point where the scan quality is acceptable. We have visited sites where scanner configuration files have in excess of 400 presets to choose from for everything from fax to invoices to laser prints. These represent a tremendous investment in time and user frustration. Contrast that with the more expensive scanners that have compensation built into the scanner and produce high-quality scans over a much wider range of originals. The higher price of the scanner gets washed out when more than even a few scans a day are going to be made.

The decision about whether to go to a service bureau or not has a pretty straightforward cost analysis: the cost of equipment purchase and direct labor costs versus the cost per page of scanning by an outside agency. Often, factors such as document security or continuous accessibility will prevent a service bureau from being used. Still, a hybrid approach can be adopted where new documents are done in-house while back-conversion of existing documents is done outside. In such an approach, however, care must be taken that indexing and file-naming conventions are well understood and executed so that the two sets of images play nicely together. The last time we calculated an exact cost break-even, it was about 8,000 documents per year. Above that, it was more cost-efficient to go in-house. But, recent reductions in scanner prices and innovative new methods on the part of service bureaus have, no doubt, changed that point.

The type of document also can influence the decision. We were once back-converting engineering drawings for an organization, which had, about a decade earlier, decided that the "boilerplate" section of the drawing should be updated. A 4" x 6" note was pasted over the old wording on about 1,000 drawings. As the papers sat in large flat files, the cement oozed out around the edges of the add-on sheet. When these were scanned, the cement gummed up the scanner rollers so badly that finally all the masters had to be photocopied before scanning. Too late we learned of a service bureau that filmed originals with a camera and then scanned aperture cards rather than scanning the paper. Such a noncontact method is ideal for worn or valuable masters.

The other half of the "what" question answer is electronic. Access to the electronic form of documents greatly reduces the operational problems of creating image data elements. Those data elements can either be left in the native form or converted to a standard image format for consistency and lower training burdens on end users. Access to the electronic form also makes it very practical to incorporate indexing schemes such as full-text retrieval, where, rather than keywords, the entire text of the document is available for searching. In environments where both electronic conversion and scanning are used, there is very little synergy required to reduce duplication of effort.

Facsimiles (faxes) fall into that unusual limbo of being inherently electronic and almost always in paper form. Many good systems now exist for the electronic management of faxes, and businesses wishing to streamline workflows and integrate imaging systems should evaluate those with high priority.

The "who" question centers on whether a single individual or group does scanning and indexing, or whether a broadly spread community is allowed to scan



and add documents to the system. As the volume of scanning builds, organizational pressure most often pushes scanning out to more and more user groups because "such a large group just does scanning!" is hard to look at on an organization chart. Whether this is the right decision is subjective, but there are more risks associated with widespread scanning due to bad scans, bad indexes, and electronically misplaced images; security, audit trail, and training costs can also have negative effects on widespread scanning. In some cases, particularly when contributors are geographically distant, a good option is to submit scans and associated images to a central group of auditors who confirm quality and naming conventions and then submit the images to the on-line system.

### **How Will Data Be Extracted from the Images?**

In order for a text retrieval system to function, it must have a body of text from which to build an index. If the document retrieval database is being built from images of documents, the images of text must somehow be converted into ASCII character representations of that text. This can be done manually by having the text keyed or by passing the images through some type of automated optical character recognition (OCR) process.

#### ***Manual Description and/or Rekeying***

Manual keying of textual data has long been the primary method for transforming paper data into electronic form. The processes for data entry from paper are well understood, and there are many firms that specialize in contracting for data entry. These firms provide scheduling, facilities, coding, and quality-control services. The costs for these services are based on the amount of data to be keyed and the desired level of quality. Charges typically are based on a per-kilocharacter amount; typical charges as of this writing start around 75 cents per thousand characters keyed. The advantage of manual keying is that it is an easy task to outsource with predictable costs and results.

#### ***OCR/ICR***

The process of optical and/or intelligent character recognition (OCR/ICR) involves the automated recognition of ASCII character data from an electronic image. If the documents to be processed exist only on paper, then the first input step for OCR is to scan the documents to produce an image file. The OCR software then processes the image through pattern recognition algorithms, producing the ASCII output. It is much easier for OCR programs to decipher typed or printed text than it is to recognize human handwriting. Even in printed texts, the variability of type fonts, sizes, and backgrounds can have a deleterious effect on the accuracy of OCR processes, as can stylistic variations such as italic or outlined type. Very small print (generally 6-point type or smaller) is more difficult to recognize. The proper recognition of formatted tabular data is also a large consideration in many applications and is not equally

supported in all software. Current programs can convert printed texts with an accuracy rate of approximately 95 percent. In most cases, the data produced by OCR will need to be proofread and edited before loading into the database. In some cases, it may turn out to be cheaper to have the data keyed manually than to begin with OCR processing. The system designer will have to decide for each application whether OCR can produce data at an acceptable error rate.

### **What Structure Will Be Applied to the Data?**

The initial evolution of electronic document handling took place in an environment where documents always existed within the same piece of software, usually a word processing system. During the 1970s, the initial evolution of electronic publishing took place, which required publishers, usually those that indexed and abstracted journals, to distribute their data on tapes to vendors of on-line information systems. Most of these publishers created their own formats for distributing their data, requiring the information vendors to create custom conversion utilities for each vendor. As data exchange became more common, standardized distribution formats, such as MARC for library data, began to appear. With the current emergence of widely networked systems, along with the emphasis on open hardware and software systems, the use of standardized document interchange formats is assuming much greater importance.

Most document retrieval systems require that some amount of structure be applied to the text of documents that will be entered into the database. The system architect must be aware of three different levels of structure: the logical document content structure, the overall document architecture, and the physical document structure. The logical document content structure is a formalization of the content of the documents themselves. In a business letter, for instance, the standard elements include the sender's name and address, the addressee's name and address, the date, and the signature, as well as the body of text. Generally speaking, a single database is made up of documents that have a similar logical content structure. In some cases, the person responsible for database preparation will have to impose and ensure that structure upon the documents to be loaded. In all cases, the database designer will have to decide what elements are required for all documents, and a procedure for coping with missing values will need to be devised.

The overall document architecture is the application of some abstract syntax to the document content. Abstract syntax here means a way of "describing data types in a machine-independent fashion." In describing data to be added to a database, for instance, there must be some method for knowing when a given data field starts, and what the content of the data field signifies. This abstract syntax is different conceptually from the physical record format. It is possible for data to be formatted using the same abstract syntax in several different physical structures. It is also possible for systems designers to develop and implement their own abstract syntax for formatting and storing data. It is more common, however, to use typical abstract syntaxes that are imposed either by the system the data are produced in or by the document retrieval software.



The physical document structure refers to the way the bytes of electronic data are arranged on whatever medium is used to store the data. This can be referred to as the "concrete syntax." While a given document architecture will often imply the use of a corresponding physical structure, it is very helpful for the system designer to think of these as two different and separate items. This is especially true in environments where data will need to be transferred among different hardware and software architectures.

There are several competing document architecture standards currently being submitted by both vendors and standards bodies. As of this writing it is unclear which (if any) of these architectures will assume increasing importance and which will fall by the wayside. The competing architectures include: standardized generalized markup language, or SGML (ISO 8879), which is beginning to be widely used in the publishing industry; the Defense Department's Computer-Aided Acquisition and Logistics Support standard (CALS, MilStd-1840-B), which incorporates the use of SGML; the open document architecture (ODA) standard, which is an emerging ISO standard for document interchange (ISO 8613); and DEC's Compound Document Architecture (CDA) format. As more vendors begin to stress interoperability, it will be increasingly important for systems implementers to ensure that the systems they work with support data exchange in standardized formats.

### *Image Architecture Standards*

There are as many ways to store bit-mapped image data as there are ways to create it. An excellent reference book is *Graphics File Formats*.<sup>3</sup> This gives a detailed look at twenty-four common image formats and a brief description of twenty-four others. In general, there are two types: raster and vector. Raster data represent a bit-mapped image in rows of image data. The resolution of the image is determined by how many bits of image data are represented per row and by how many rows there are. This information is either uncompressed or compressed via a number of lossless techniques such as CCITT Group IV or lossy techniques such as JPEG.

### **Where Will the Image Be Stored?**

One of the more arguable decisions in the design of an imaging system is where to store the images.

### *Integrated Storage*

One school subscribes to placing the image objects in the actual database schema. The argument is that this improves synchronization of the image and metadata about the image. The argument against is potentially massive databases, with unknown impact on backups and access times.

### Separate Storage

The other school argues that images should be stored separately from the database and use pointers to the image data. The advantage of reduced database size, however, is offset by the "double commit" process of updating image data separately from the associated metadata. Another drawback of that method is a reduced robustness of the imaging system in case of system failure. Because pointers are on a separate medium, loss of these pointers or of synchronization with the image content will result in a total loss of access to the imaging data.

### Media Selection

Most people considering imaging systems seem to assume automatically that optical storage, particularly jukeboxes, is the required storage medium for imaging systems. This is far from true and may be the last thing an end user wants for his or her imaging system. Unless the image database is expected to grow beyond about 10GB, the system developer should strongly consider the option of using simple magnetic hard disk storage. Systems that have a high rate of change documents may want to push even larger magnetic options. Large disk farms are certainly manageable but issues of backup and recovery become greater concerns.

Before selecting WORM (Write-Once, Read-Many) drives, the feasibility of erasable-optical (EO) and magneto-optical (MO) drives, both of which are erasable, should be considered. Highly static data and "forever" archive requirements will move the user toward WORM storage. Any choice for optical involves high media cost, potentially difficult backup and recovery strategies, and often closed system architectures. Disk swapping times in seconds to tens of seconds can impact system performance at the user level. Combination technologies, which mix magnetic and optical disks such as Epoch, can eliminate much of the system integration headaches of stand-alone optical storage systems and reduce or eliminate the user's negative perception of system latency.

---

## 11.5 QUALITY CONTROL

---

A last consideration in the design of the imaging system is adequateness of the database to support quality-control procedures. One significant claim of image-based document management systems is that of "no lost documents." This claim is true in a well-administered system, but only if the document gets in correctly in the first place. The control of the capture, naming, and indexing processes is absolutely critical. Whereas misfiled paper can be found based on other documents that were in use, which drawers were opened, and so forth, incorrectly naming and indexing an electronic image can randomly place documents anywhere within millions and millions of pages, making them statistically lost forever. Or, incorrectly incrementing an index at the beginning of the day can leave thousands of



documents "off by one" at the end of the day. This, added to an optical storage environment, can be a time-consuming and costly mistake.

The most critical data are input using a double-blind method with two operators: The first operator inputs the document ID, which is read from the first page, and the number of pages from the title bar. After scanning, a checker is presented with the image of the first and last pages of the document and the electronic page count. The quality of the scan is evaluated. The second operator checks the first and last page title bars to match page count and types in the document ID read from the image. The checking software matches the initial ID typed in during scanning with the ID read from the document. After these checks are passed, the document is available for release into the system. When batch scanning is taking place, care must be exercised that up-issues of documents are not mishandled while the current version is out for scanning. Chances are that the existing paper environment that is being replaced has a very formal, if not fast, method for controlling the release of changes to documents. Relaxing those standards with the advent of electronic imaging is probably not in the best interest of the organization. It does offer significant opportunity to gather statistics on usage and users and provide a framework for intelligent decision making in further modifying the workflow of a business or organization.

---

## 11.6 CONCLUSION: IMAGE ENABLING EXISTING INSTALLATIONS

---

Perhaps the single biggest challenge facing IS departments in the 1990s will be how to add imaging to existing systems. The "back up the truck" approach to conversion, where all the existing hardware and software are loaded up and driven away while all the new hardware and software are taken off and installed, is just not an option for most businesses. The cost in equipment, software, data conversion, and retraining is beyond budgetary limits in times of intense global competition and worldwide recessionary trends. This is true even in some cases where the payback in future cost avoidance and better cycle times is easy to see. Management has to allocate limited dollars to all areas of the business, and extraordinary expenses in IS often get cut proportionally with other business entities. So, the question remains, how does the IS manager, with perhaps half the money he or she needs for a totally new system, implement business-changing methods of handling critical information? While the answer varies from site to site, some general principles apply.

A first consideration for "the biggest bang for the smallest buck" item is usually in the area of networking. A principal benefit of imaging in the enterprise arena is that workers who previously had access to documents only through paper now have that access electronically. Network upgrades that allow fast, economical transfer of large binary objects to all areas of the business will have lasting payback. Connections that allow heterogeneous computer environments to share information, both text and binary, inside and outside the company, are the single most

important aspect in creating a paradigm shift in the way companies design, manufacture, sell, and support products and services. (The topic of networking is treated elsewhere in this book.)

A second consideration in adding imaging to an existing environment is salvaging hardware, facilitated by the fact that the days of proprietary imaging workstations are at an end. The challenge now is to select software that will allow existing computers, often from different manufacturers with different operating systems, to access image data over the newly upgraded network. Two methods are used to address this: terminal emulation and multiplatform software. The first method involves adding terminal emulation software to existing DOS, Macintosh, UNIX, or other computer environments, with all users then actually running from one or more dedicated servers. The advantage is that all data and the "core" image processing is done centrally, where it can be better monitored and controlled. As the user base grows, however, that advantage also becomes the disadvantage, because all user functions, including pan/zoom/print are done on the image server, with a large amount of network traffic supporting the many users accessing that host.

The second method involves a more client/server approach where image viewing and manipulation software is installed on each user's computer and the image objects are sent in compressed form from the storage device to user, where decompression/pan/zoom/print are supported. Image metadata, index, and retrieval information can either be located and accessed on the central server or distributed around the network as well. Some systems allow both methods to be implemented successfully, and the former approach is used for departmental data, where a limited, controlled number of users access image and indexing information on a UNIX workstation via X-emulation software from PCs and Macintoshes. The latter is the case for enterprise-wide applications. An example is in an electronic printroom shared among different organizational units. In a particular case, five regional "printroom" servers are accessed by over 3,000 user clients. Compressed image data are sent to the user where the decompression/viewing is done. A customizable API (application program interface) allows the same viewing software to be used for both applications.

Image enabling an existing system requires the addition of information about the images to a current database. As discussed above, image metadata are fundamentally different from other data in the system. For the most part, they contain no information that is of primary interest to the end user. They fail the relational test given at the beginning of the chapter in that they are not perceived by the user as a table, and their function is not to generate another table. Once again, there is no "one size fits all" solution to adding the image metadata to a database, but it should be a major area of interest in the design and implementation. The creation of the metadata and their inclusion in the database should be as automated as possible. Ideally, all metadata required to select an image exist in the current database, with the exception of where they are physically located in the file system. Information such as date, last access, revision level, document ID numbers, customer reference number, associated PO, or part numbers, which are already captured, are enough to narrow a search to a single image document or at least a set of image documents.



Often, such information can be captured via selective OCR, or ICR as it is sometimes called, as part of the scanning process.

Once the network, hardware, software, and metadata are in place, and the images are captured and indexed in a file system, the last element of image enabling is the design and implementation of the user interface to proceed from narrowing a search to identify a single image to the retrieval and viewing of the desired image. The simplest case, an archival or document distribution system, requires only that a single "button" be added with a descriptive title; "view image" leaps to mind. Transaction-oriented data, especially that where images are annotated as part of a review or sign-off process, require a more robust user interface and document control methodology.

---

## NOTES

---

- <sup>1</sup> Belasco, J., *Teaching the Elephant to Dance: Empowering Change in Your Organization* (New York: Crown Publishers), 1990.
- <sup>2</sup> Peters, T., and R. Waterman, *In Search of Excellence: Lessons from America's Best-Run Companies* (New York: HarperCollins), 1982.
- <sup>3</sup> Kay, D., and J. Levine, *Graphics File Formats* (Blue Ridge Summit, PA: Windcrest/McGraw-Hill), 1992.

---

# 14

---

## Tools and Toolkits for Network Imaging and Workflow Systems

---

A. J. Wand

Plexus Corporation

### Editor's Notes

*The preceding chapter dealt with image management development software, seen from the perspective of image enabling an existing system. The focus in this chapter is on the application of such development tools in a context of downsizing, and it shows how the use of various commercially available tools can facilitate and reduce the costs of the integration of imaging in a redesigned workflow.*

---

### 14.1 INTRODUCTION

---

You can still hear the cries of yesteryear reverberating through the data center: "What's good for the mainframe is good for the company." Host-based data processing was a powerful means of automating transactions in yesterday's paper-based world, allowing data to be managed and worked on in a central location. But as important as this on-line transaction processing was in making businesses more efficient, there was still a big issue relative to data processing: Some 95 percent of the corporation's data remained on paper, long after the mainframe transaction had been completed.

The downsizing movement has brought a number of options for processing corporate data in a local area network environment, and imaging is an enabling technology for companies that are automating work on vital business information.



The chief virtue of imaging is that it makes available on the desktop approximately 90 percent of the data that do not reside in the traditional data processing application, allowing people to work faster and more effectively. "Compound data processing" involves more than just alphanumeric data, allowing access to *all* the data relevant to the transaction, regardless if they are paper-based or audio-based.

---

## 14.2 THE NIWS DIFFERENCE

---

### The Lay of the LAN—Open Client/Server Applications

A primary difference of network imaging and workflow systems (NIWS) is a distributed LAN-based architecture, which is new technological territory for many MIS organizations. Traditional host-based data processing systems were monolithic—the mainframe provided every element of the application. The system required little or no integration, and issues such as system compatibility and communication protocol were nonexistent.

Client/server applications represent a fundamental shift in MIS viewpoint; whereas mainframe applications are coded such that data storage, computation, and presentation occur in the same machine space, client/server applications require that the developer break the application into two pieces: The client handles the application logic and presentation of data queried from the server.

Today's LAN-based systems offer a lower-cost alternative to mainframe and minicomputer applications, distributing the overall computing workload across a variety of lower-cost (for both hardware and software) machines. The process of distributing applications is not, however, a free ride; there are costs associated with learning new architectures and allowing heterogeneous machinery to interact in enterprise applications.

### Compound Data over the Network

Mainframe applications with terminal-based user interfaces traditionally do not manage compound data-like images. On-line transaction processing originated in a text-only environment, with host systems handling critical alphanumeric data that required state-of-the-art data management and security. These proprietary systems have a high cost-per-transaction, and some feel that mainframe cycles are wasted on image processing, and that smaller, cheaper LAN systems are better for imaging applications.

Local area networks, too, were originally designed to carry relatively small amounts of data (also typically alphanumeric) and not large bit maps. Ethernet has a bandwidth of 10MBps, with a maximum theoretical throughput of approximately 3MBps. A distributed application using uncompressed 300 dpi images will quickly consume this bandwidth. LAN-based systems allow processing to be distributed, with local processing available to handle database querying, image decompression, and other local operations to boost the effective throughput of information. In

addition, overall throughput of currently popular and affordable network topologies can be compensated for by using multiple networks to service workstations if volume is high. UNIX servers can handle a large number of networks simultaneously, in addition to storage devices and other peripherals required.

### **Sophisticated Work Processes**

Older host-based applications inherently embedded all of the system facilities in a single environment, where the mainframe handles all aspects of the application user interface, data storage, communications, workflow routing, and so forth. All of these facilities are embedded in host systems, rather than a number of independent activities that communicate across networks and literally move work from one to the next. With client/server workflow applications, workflow is no longer deeply embedded in a monolithic application and difficult to change. Autonomous applications process information in a complex, orchestrated workflow that may be based originally on a manual paper workflow.

---

## **14.3 TECHNOLOGIES THAT MAKE A DIFFERENCE**

---

### **Imaging as a Basic Technology for Work Process Automation (WPA)**

Although once strictly the domain of specialized, high-cost systems, imaging technology has become a widely accepted commodity that is integrated with standard application functions. But imaging is no longer an end unto itself; it has evolved from "an application" (insurance claims processing, for instance) to a generalized application *service* that provides richer information in many applications. Imaging must be considered as a basic technology in the overall scheme of automating business processes, much like the windowing or network software on which the system is based.

### **Optical Storage**

High capacity and low cost per gigabyte (compared to paper filing cabinets or mainframe DASD) make optical storage attractive. Assuming 6GB capacity platters and compressed images 60KB in size, each platter can store approximately 100,000 pages and provide much faster (on-line) access. Companies can literally replace buildings with optical jukeboxes, saving precious time and money in the process.

Another benefit of optical storage is endurance. When critical paper data are imaged, they need to be available for a period much longer than the life of dismountable magnetic media (i.e., tape). Optical storage provides this durability, along with the security of a write-once medium.



## Graphical Environments

GUIs provide the consistent and flexible interface required for dealing with richer data. Richer, more sophisticated data present additional manipulation and management requirements that are not present in text-based applications, and current graphical software environments that support bit-mapped displays, multitasking, sophisticated windowing, powerful interapplication communication, and other features are an obvious advantage in compound data applications. Not only are applications in this environment more effective for this type of processing, but the development tools available to build those applications are better as well.

## Imaging and the Database

Many imaging systems store alphanumeric data and image data in separate facilities, requiring coordination of the two subsystems to maintain data integrity. The image server may consist of a standard relational (or proprietary) database to store the alphanumeric data, and a proprietary image server manages the optical hardware devices. The servers communicate using remote procedure calls or network messages.

Although a loosely coupled approach allows more database independence, the system can experience integrity problems if it is not of robust design. For example, if an optical server fails, data records may be stored in the application database without the associated images written to optical if the system is unable to roll back this partial transaction. Keeping distributed systems "in synch" is of primary importance, as images stored on optical are a crucial portion of the application data.

Another approach is a tightly coupled one, where images are stored in a relational database itself. The capability to store BLOBs (binary large objects) is becoming more prevalent among the major relational database management systems (RDBMS) vendors. These systems manage large data objects (e.g., images) as data that are as basic to the application as the alphanumeric information contained in a database record. Relational databases are designed to manage large amounts of related data and can provide all the integrity, concurrency, availability, and security for both image and other related alphanumeric data. In addition, allowing the DBMS to manage the archive and logging of images is significantly easier than developing the archive as a separate application service.

## Optical Character Recognition

Optical character recognition (OCR), long a boon to data entry operations, is an important technology in imaging, particularly for forms processing applications. OCR allows autoindexing, full-text search, and other capabilities that make it easier to manage and organize scanned data with less user interaction.

OCR is used in applications—both for data entry and indexing of forms, as well as text recognition—that often are used for full-text indexing the contents of documents. OCR's effectiveness depends on a variety of factors, including docu-

ment quality, quality checking required for recognized characters, and the strengths of the specific character recognition product used.

Effectiveness is usually calculated in such a way that recognized characters that are guessed incorrectly carry a much greater penalty than characters that are unknown. This is to say that OCR provides a benefit only if a high percentage of characters can be recognized, and only if these characters are recognized correctly almost all of the time. Some implementers find that the cost of OCR is actually greater than a manual process, while some have "recognized" immense gains in productivity. Expertise with these specialized types of applications is important to making them successful.

---

## 14.4 WORK PROCESS AUTOMATION AND BUSINESS REENGINEERING

---

### Ad Hoc versus Structured

Workflow is the automatic routing of information between users so that they can process the data in some fashion and add value in performing business functions. This includes both the actual routing of the data, as well as the effective management of that process. Workflow can be divided into two broad categories: information-based and transaction-based.

#### *Information-Based Workflow*

Information-based workflow is the ad hoc routing of data between processing activities. This work process requires that the sender direct information explicitly to another user or activity, as in the case of electronic mail. In simple ad hoc workflow, users might process information on their desktops and route it as necessary using a basic technology such as e-mail. This type of workflow is essentially manual, where routing rules and decisions exist through the intelligence of the user and are not part of the automated system. The system simply provides the transport layer (how to send the data), but provides no other help in automating the *process* (when and to whom to send it). Information sharing systems (a.k.a. groupware) support worksteps that are ad hoc in nature, but do not automate work processes. The term workflow in these systems is more loosely defined and ad hoc than other process-based systems, where work can follow predefined paths for processing. Transactions don't need to be monitored and recovered by the system, because the users interact and perform this function in the course of their work.

One example of ad hoc workflow (that is not groupware) looks like an extended storage-and-retrieval application, where routing is automated, but processes are not. Consider a medical records application where a doctor's initial retrieval of information may be triggered by a patient's arrival, a phone call, (e-)mail message, whim, and so forth. When the doctor is finished processing the patient information,



he or she might determine that it needs to be routed to a certain specialist for more scrutiny. The routing is performed by the system and replaces a paper-based process where the doctor makes the same decision and routes the patient's paper files to another participant in the process. The doctor could alternatively send an electronic message to the specialist instructing him or her to retrieve the patient's information for processing.

In this example, transporting work electronically provides benefits for information sharing and communication, and to some extent flow management, although there may be no structured work process to be automated and evolved. All processing is essentially "exception processing" that is not predetermined. Workflow solutions can provide more functionality when there is a set of steps to be automated above and beyond the sheer physical routing.

In applications that involve a structured process, the flow of work can be automated beyond providing a transport mechanism, where both the routing and the work process are automated by the system. Structured processes provide an opportunity for transaction-based workflow and work process automation to add considerably more value, as computer systems are exceptionally good at automating tasks that are structured and repetitive.

#### ***Transaction-Based Workflow***

Transaction-based workflow involves an active process of routing data, where rules and routes for the information flow are predefined. In this case, workflow should not only automate transactions, but should also make it possible to deal with exceptions and automate *them*, thereby providing the flexibility to adapt to exceptions that become the rule. This type of workflow is particularly effective when used in combination with imaging for compound data processing.

Transaction-based workflow requires an engine to monitor and manage workflow processes and information. Database management systems are good foundations for these sorts of tasks, where e-mail-based systems may not provide the appropriate transaction or state model for work process automation. E-mail may serve as a transport mechanism for data, where the state information is stored in a transaction-based system.

Work process automation systems provide another dimension to the solution in that they transport not only data, but allow organizations to automate and manage the underlying work process.

#### ***Interactive Workflow Management***

Ideally, the rules for workflow routing are defined separately from the image processing applications in client workstations so that as the business environment changes, the procedures (i.e., routing rules) defined for processing information can be adapted with minimum effort. The embedded workflow found in image processing systems uses hard-coded routing instructions within desktop applications, requiring code changes in each and every desktop application that participates in the workflow.

Client/server workflow is a relative newcomer to the compound data processing scene and represents an important development in the integration of imaging with traditional applications. Clients in the workflow communicate with the workflow server according to an application programming interface (API), using callable libraries for the various workflow functions. So rather than a client application routing data to the next processing activity using explicit routing instructions, it simply "asks" the workflow server to route the data to the next process. The workflow server understands the routing rules and puts the data in the appropriate queue, where they wait until they are requested by the next application that needs to process them. When workflow is provided as a network service to desktop applications in this manner, users can easily reengineer their business processes for maximum efficiency and effectiveness and dramatically reduce the expense and complexity of adapting as a business evolves and work processes change.

### Using Workflow to Integrate Compound Data

The power of the client/server workflow model can be applied to integrate existing applications with image. The workflow server communicates with various systems, and routes data to applications either by *reference* or by *value*. *Data by reference* means that the workflow client receives a message that points to the data's location, rather than to the actual data. The client application must then be able to retrieve the data from the application/image server with its own access routines. *Data by value* is when the workflow server can fetch and route the application data (e.g., alphanumeric, image, other binary) directly to the client for processing. Note that image and workflow servers may or may not be on the same physical machine.

---

## 14.5 WPA TOOLS

---

Advances in computer technology, application software, and communications have contributed to the proliferation of PC-based workgroups within businesses. Workgroup members communicate and share information over local area networks; however, the actual workflow, the job of moving information from one processing step to the next, remains a task that is handled either manually or electronically by a complex computer program.

### Choosing Your Weapons: The Pro-Am

Whether application development is provided by internal programmers (MIS or department) or by outside value-added resellers (VARs) and systems integrators, robust flexible tools are a requirement. Fourth-generation languages often provide the right combination of flexibility, power, and ease of use for the professional developer of work process automation systems. These development tools feature a comfortable combination of high-level functionality (user interface building, high-level language, integrated functions, etc.) with the flexibility to use lower-



level facilities (e.g., granular control of application behavior, integration with C-level services) when necessary. The limitation is that applications and facilities developed using these tools can be changed only by the professional programmer.

Robust application development tools should allow the professional developer (e.g., MIS) to build a set of activities (applications) that allows a department to process its work. In a forms-based application, for example, the forms that are processed (and therefore the applications used to process them) will probably not be changed a great deal over short periods of time. The workflow between the steps, however, may need to change significantly. Once those individual applications are complete, the separation of the workflow from the work processing application allows the department to determine the proper workflow to connect those individual activities. Although MIS, for example, may provide basic application development, it is the department staff that understands, and therefore should be able to modify, the work processes.

Department-level "developers" require different tools that allow them to express and implement their knowledge of the business procedure without system-level programming. Today's graphical tools can provide such a facility for nonprogrammers to control work processes without modifying applications, provided that workflow is managed outside of the imaging application. Once the applications are developed for specific processing steps and activities, these tools typically provide an environment where "developers" manipulate icons that represent the processing activities (e.g., scan, approval) and control routing behavior by connecting the activities to form workflow paths. Activities may represent hundreds of workers all performing the same task. The developer can modify the rules that govern workflow routing decisions without modifying the processing applications themselves.

### **Criteria for Toolkits**

Tools and toolkits come in many different flavors and colors, promoting different orientations, features, and styles for developing applications. Even with these differences, a number of the issues that toolkits help users to manage are universal ones: for example, providing ease-of-use without sufficient flexibility. The following are some of the basic features and capabilities to look for when selecting work process automation toolkits and systems.

### ***Development Environment for Compound Data Applications***

The toolkit should be integrated for the specific purpose of imaging and compound data management and should take advantage of operating environment-specific features as much as possible. Key features include graphical user interface builder and a fourth-generation language, preferably in an environment that was designed to build compound data-based programs. This type of integrated environment provides a more robust and efficient solution for sophisticated development than does cobbling together a solution with a raft of components.

Peripheral integration is a task that doesn't provide much value for application developers.

### ***High-Level Interface to an RDBMS***

SQL, for instance, is a standard language that provides the programmability to support any type of data model. Programming languages may use a folder construct, and although it provides a high-level interface, the data model is somewhat limited.

### ***Transparent Support for Optical Storage***

Access to and management of optical storage of application data should be integrated with the storage of alphanumeric data for maximum benefit.

### ***Programmability and Flexibility***

Applications should conform to the demands of the user, and not vice versa. Unfortunately, toolkit flexibility is often lost at the hands of ease-of-use, though this is not to say that these two objectives are incompatible. If a system does not provide configurability to fulfill the application's requirements, then the full value of computing systems may not be realized.

### ***Changing Workflow***

The capability to change the workflow "on the fly" without modifying or affecting applications allows the fastest and easiest possible alteration in business policies and procedures.

### ***Control of Workflow***

When users can monitor, analyze, and test changes to workflow, they can adapt to the ever-changing world as safely as possible.

### ***Graphical Workflow Builder***

Visual workflow programming is significantly easier than developing scripts in a 4GL, which allows nonprogrammer "experts" to modify systems. To be effective, such workflow builders should use a database as a robust workflow engine, because business processes imply transactions, and DBMSs are excellent at managing this sort of workflow metainformation. However, simplicity of use requires that such builders do not involve database coding to program a workflow; the database is simply an engine for the workflow management software. Consequently, workflow developers and administrators should only need to understand the work processes, not develop the underlying database code for managing workflow.



### Workflow Management Tools

Workflow is not static; therefore, the benefits garnered from process redesign are only as good as the tools provided to implement and change the flow. Workflow redesign tools should provide benefits in several major areas concerning the development of workflow rules, and how, when, and by whom they can be changed.

Optimally, tools should make it easy for the "experts" to set up and modify workflow. In this case, "experts" refers to the organization that does the work and implements the workflow (e.g., department analysts), rather than the implementers of the base technology (e.g., MIS). If technical programmers are required to change the flow, the benefits of fast and flexible process redesign may be lost.

A methodology and validation for the workflow should also be provided so that "experts" can examine the effects of redesign. This includes feedback on system performance to help determine what adjustments should be made and a means to test the effects of a change to the flow.

Even though individual applications that perform work functions may not change very much after implementation, the processing flow will. Users will discover that they modeled existing (paper-based) workflow properly, but that the flow can benefit from modification. Ideally, workflow can be changed easily over time using the proper tools, without requiring the recoding of desktop applications. Users will likely find that adjusting workflow for maximum efficiency is an ongoing process (see Figure 14.1). Consequently, there will be a need for tools that provide easy ways to describe present requirements and address new ones.

### Scripting versus Graphical Tools

Workflow scripting is one common method that imaging vendors provide to flow enable imaging applications. Scripting provides a high-level way to define routing rules without having to understand the lower-level implementation of work queues, transports, database access, and so forth. Scripting, however, is still like high-level

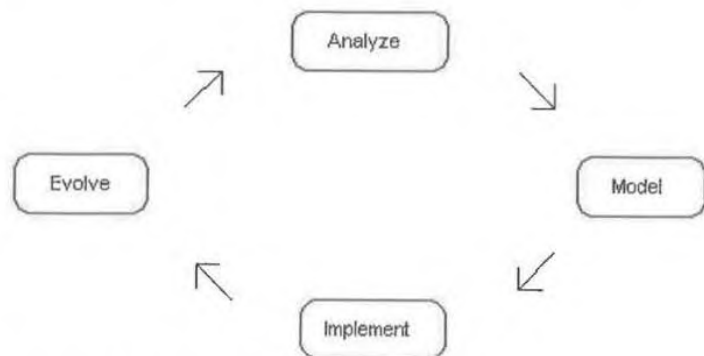


Figure 14.1 The process improvement cycle

programming; its syntax is strict, and some knowledge about transactions and queues (i.e., programming issues) is still required.

Ideally, work process automation tools allow user analysts to "draw" workflow maps in a graphical environment, where the different processing steps are represented by icons. The developer must understand only the work process steps and activities and need not concern himself or herself with structured programming or with the specific issues of underlying workflow data management. Once the new workflow is completed using a graphical tool, all database structures necessary to implement the workflow are automatically created without need for coding.

Another important capability of workflow tools allows administrators to change workflow easily and interactively. Workflow may be developed in various ways, usually using either a workflow script or a graphical tool. The key issue is that if routing rules are embedded in the desktop applications, then the workflow cannot be changed without changing each application that participates in the workflow.

### **Examples of Workflow Tool Capabilities**

There are various types of tools on the market, with different capabilities. Choosing the best tool depends on the applications, the scope of the project, and the strength of the support staff. The most common characteristics are outlined below.

#### ***Interactive Tools***

Interactive tools allow organizations to react quickly to procedure changes by modifying the workflow and adding new processing steps outside of the existing applications. Tools should allow changes to the flow while work is in progress, without disturbing users or the work flowing through the system. Once the new workflow is complete, whether by graphical tool or scripting, all database structures needed to implement the workflow should automatically be created without additional lower-level coding. In this way, an interactive workflow system can adapt to handle exceptions and accommodate changes in routine business operations without recoding, recompiling, or retesting individual applications that are already deployed.

#### ***Monitor Status of Processing Activities***

Another useful workflow tool is one that displays the status of the work (e.g., number of users logged into the activity, the priority of the activity) that is flowing through the system for each work processing activity. Some information can be displayed graphically, rather than in a textual format, providing direct visual feedback about where the work is and how it is flowing. This type of tool can be helpful in monitoring and managing workflow performance, allowing managers to use real-time information to make decisions about day-to-day work processing.



***Test Workflow before Going On-Line***

Just as an automobile manufacturer might simulate the design and performance of a new sports car before putting the model into production, complex workflow requires a means to analyze system performance *before* it is put into use. Simulation tools provide a "sneak preview" of changes in workflow without processing real data, running a virtual model based on the rules and routing information defined in the system workflow map. Work processing activities are simulated by selecting and varying expected response parameters (such as delay time, output state of document, etc.), thereby modeling system performance before using live data. The system can then be adjusted off line to reach an optimal configuration before putting a new flow into practice.

***Management Reports on Organizational Efficiency***

Workflow management systems should also be able to track the history of work through the system for true management information. System data, including information about work items such as date and time of day, activities performed by users, and so forth, should be logged in a database. System administrators can then analyze an organization's efficiency by creating custom reports from the logged workflow metainformation using a standard DBMS report writer tool in much the same way these tools generate reports for application data.

***Control and Integrity of Workflow Data***

One of the issues of concern with workflow systems, as with other data management applications, is how and where the rules and state information are stored. Process integrity in transaction-based workflow requires that control and management of the workflow be from a single point, most likely using the services of a database management system. Work in the system may carry its own routing information, though a central workflow database is still needed to track the actual flow. If desktop-based scripts control the workflow, transactions and consistent flow are not assured, and control over the data is lost. Process integrity allows industrial-strength applications with sophisticated workflows that interact with other complex systems beyond the scope of the local flow.

---

**14.6 FACTORS IN DESIGNING WORKFLOW SYSTEMS**

---

Because workflow represents a physical process, workflow design must follow specific rules that reflect practical constraints.

**Processing Activities**

The system designer(s) must decide how to divide the application work into different workflow activities to maintain overall application modularity. Existing

applications should be divided into separate processing activities to allow flexible routing between different application steps. Newly developed custom activities would require only an interface to the workflow manager.

### Input Devices

Inputs into the workflow system might include optical scanners, tape drives, gateways to and from another system, or keyed data entry input from a local application. Software applications associated with these input devices are written to process data and move them from the input device into the imaging and workflow system.

Depending on your requirement, you may choose either a centralized or decentralized approach to document image scanning. Some applications are best served by the "few plow horses" approach (high-speed scanner[s] attached to the server), while others are better suited to "hundreds of rabbits" (desktop scanners attached to a number of workstations) to perform the work in a more distributed fashion. The approach you take depends on how documents enter and are processed in the system, either through a central source (e.g., the mail room) or distributed across a number of scanning workstations.

### Input/Output to Workflow Activities

An *input interface activity* takes data controlled by the workflow system and writes these data into a form usable by an existing activity and even executes a command known by the existing application to tell it about the new data.

An *output interface activity* either monitors the output data or is triggered by some command from the existing application. The activity retrieves these data from the existing application and prepares them for use by workflow activities downstream.

### Data Access Considerations

Whether the workflow system can access data by reference and/or by value affects system attributes. If the system routes data by reference, then all access methods and drivers (e.g., database, file systems) must be present in and performed by the applications. The workflow manager tells the workstation where the data are located, and the software on the workstation retrieves the work. If the workflow manager can retrieve data for activities directly, then access methods for data need not be present in the workstation.

### Multiple Server Considerations

When developing systems for future expansion, it is important to consider workflow management support distributed across multiple servers. In some nondistributed environments, gateways can manage the flow of information between single



servers, although administration and data management must be handled at each server.

### **Processing Time for Each Activity**

In planning the layout of workflow maps, it is important to calculate the duration of each activity. Appropriate numbers of users and resources must be available for each activity to prevent data flow bottlenecks, during which processing slows down to unacceptable levels. Bottlenecks also require more overall disk capacity for storage of the "slowed down" work; also, be sure that *destructor* activities that take work out of the system (for example, an archive activity) are present in the flow so that disk storage requirements are controllable.

### **Processing Hardware Requirements**

The hardware requirements for a workflow manager are driven by the software environment and its system capacity requirements. Server hardware must be able to support the operating system activity, network communications, and the workflow database. (Note that some databases use the CPU more intensively than others.) System memory requirements are mandated by performance, the number of applications running, and the number of users anticipated on the system. The available memory must be large enough to prevent system swapping during production. Minimum memory configurations for a small UNIX system may be 16MB, though a typical small production configuration might require at least 32MB. In high-performance environments, additional RAM for caching critical database indexes could help.

### **Workflow Mapping Features**

A map represents the procedures within an enterprise. It contains a graphical diagram of each step (or activity) in a procedure with connections between steps to show the possible routes between activities. Maps in work process automation tools have several features of interest.

#### ***Properties of Maps and Submaps***

Maps (and submaps) allow the graphical workflow developer to put together small, easily tested maps and then use them in higher-level maps over and over again. Maps are like individual activities in that they have defined inputs and outputs, and once a map has been created to perform a function, it should not be necessary to recreate the map to perform the same function in a different context, even using a submap multiple times within the same map. Maps should also be able to reference themselves for recursive workflow.

Workflow systems should support a repository for all activities, including submaps, which have been created for the system and promote the reuse of processing activities and submaps.

### *Routing Behavior*

Businesses have a number of procedures in place at any given time, and a workflow system must be able to accommodate new maps or procedures easily. Some procedures require that a processing step be repeated until reaching some (successful) conclusion. Another desirable feature is to associate a deadline with the event so that administrators are alerted if work "gets caught in a loop."

Some procedures occur over protracted periods of time, so maps should be able to change at any time while work is in process without disrupting work already in the flow. This guarantees that a procedure is followed when a user initiates the procedure, and that work follows the routes it was intended to follow. Users should be able to create maps, route work, and then modify the flow for a slightly different purpose, knowing that any work in progress will reach its intended destination.

Workflow systems should be able to route data based either on the "state" of the data leaving an application or using external "rule engines," which interpret user-defined rules after an application has completed its operations. Rules may be invoked as a result of a user's selection about the state of the work or based on a data value in the work itself.

### *Activities*

Exception activities and global resource activities should not require explicit input routes. Exception activities do not execute automatically, but accept work that requires "hands-on" processing until a system administrator can take appropriate action.

Activities may be assigned a default priority, with the ability to change priority based on the amount of work queued for the activity. System administrators should also be able to change priorities.

Some activities need to participate in workflow in more than one map. These "global" activities (e.g., print, fax) may be referenced in several maps for different types of work and may need to run unattended.

### *Priorities of Work*

Work is usually ordered in queues by priorities. The priority of a piece of work may need to be changed by a processing application, by a system administrator, or by the workflow system itself, based on factors such as the age of the work item, an irate phone caller, and so forth. In addition, users may need to bypass the normal work queue and explicitly select an item of work for accelerated processing.

Workflow systems may also allow users to set deadlines for completion of a procedure, elevating the work's priority as the deadline approaches to expedite the



operation. Alarms should also be able to send a message back to a user or administrator as a deadline approaches.

---

## 14.7 WORKFLOW AUTOMATION AND IMAGING TOOLS

---

WPA is an integrating, not an isolating process, so the resulting architecture should be compatible with existing corporate computing architectures and open to embracing (nonimage) applications. Although very few people can predict the future, developers must try to plan systems that are both compatible today and extensible enough tomorrow to adapt to changing business and computing environments.

### Flexibility and Ease-of-Use

Graphical tools (and to some extent, scripting languages) provide the facility for system users or administrators (i.e., not application programmers) to modify workflow for maximum efficiency. The developer in this case does not need to know the underlying programmatic concepts of queuing using a relational data model, what work information to track and how, and other lower-level concepts of workflow automation. Graphical tools allow workflow development at a higher level, so that the developer needs to understand only the work process (i.e., what needs to be accomplished), but not how to accomplish it programmatically. Therefore, applications can be developed faster, since application programmers will write code to handle the actual work process, and this code will be modularized using a workflow API. Department staff can then implement and change the flow without writing additional code.

### Transparency

Imaging tools should allow developers to concern themselves with concepts and functionality at a sufficiently high level. An example of this is seen in the properties of SQL (Structured Query Language), a high-level language for managing relational databases that allows developers to declare *what* to do with the data, but not *how* to do it—the database “understands” the underlying mechanisms for sorting, retrieving, and manipulating the data. This concept can be extended to many (other) aspects of an imaging and workflow toolset, where the programmer is shielded from the system-level mechanisms required, and only needs to be concerned with the higher-level needs of the application.

Data manipulation, including interfaces to imaging peripherals, is one of the areas where tools should add value. Interfaces should be independent of specific hardware devices and data types, and image manipulation functions should allow the developer to concentrate on solving more abstract application requirements, rather than low-level technical ones.

Ideally, imaging tools make it easy to choose the right *data storage* medium for an application. Storage should be supported in a (relatively) transparent manner, so that developers don't have to understand all of the underlying mechanisms for managing storage on optical or other large-scale media.

The implementation of optical storage for image data is different from the magnetic-based one due to the unique characteristics of WORM optical storage, particularly of its write-once and dismountable nature. Any implementation that assumes the constant availability and random access characteristics of magnetic disks will collapse when applied to an optical environment because of the amount of time required to fault a platter, even in a jukebox environment. The overall architecture must allow for dismountability (i.e., the database data may not be available at the moment they are requested).

When an application references an image, say for retrieval, it should not need to determine the exact location of the data. If the optical platter holding the data is mounted in a drive, it is accessed, and the operation is completed. If the platter is in a jukebox but not mounted, the software determines if there is an available drive in the same jukebox. If a drive is available, robotic commands are issued to spin down and remove the platter in the desired drive, and then retrieve and spin up the needed platter. If a drive is not available, the request is queued, subject to time-out limits set by the application. If the required platter is not in a jukebox, similar results can be effected through interaction with an operator (i.e., a request to import the platter into the jukebox). The use of optical storage should be functionally transparent to the application, which does not need to be aware of the location of specific images or manage the drives, issue mount requests, and so forth.

Imaging tools should also allow for appropriate media organization for optical storage. Applications do not have to understand the issues of two-sided removable media when only one side may be accessed at a time. In addition, optical must appear as an open-ended medium that can be extended indefinitely to take advantage of the large-scale storage. Since the time to exchange platters in an optical drive is high compared to the time to access the data, the organization of data across multiple platters is important. Data may be organized either sequentially or clustered. Sequential storage is straightforward: Data are added in order, much as on a tape device. Clustered organization is important when active data reside on many volumes. Optical clustering is a means for grouping logically related images so that related images can be retrieved together with a minimum of optical platter accesses. For example, typically a user will want to view all documents associated with a given entity together. If the images are spread over a number of volumes, the retrieval time could be several minutes. Even worse, the aggregate of multiple requests to the jukebox may swamp it, resulting in delays of tens of minutes, even hours. But if all the images are located on the same optical platter, there will be no more than one platter exchange and very reasonable response system-wide.

The dismountability and write-once (or erasable) nature of optical media introduce a number of issues for an imaging system that preferably are handled transparently. Optical disk volumes must be labeled and location tracked without requiring the user to do all of the management of the media. Optical disk volumes



must be scheduled and mounted in drives for access as needed, including controlling the optical jukebox robotics. Platters may be moved into or out of jukeboxes with no effect on the application software.

### **Openness and Adherence to Standards**

Today's corporate information systems are true technological melting pots, consisting of heterogeneous systems that must communicate to allow a diverse organization to function. Openness (usually) means that a wide range of custom and commercially available components can add value to the system as a whole at competitive prices. The solution is not bound to proprietary hardware and often supports a number of different vendors' hardware platforms, networks, peripheral devices, and the like.

Standards are necessary to ensure that systems will be interoperable and easy to maintain. Images stored in a standard format (e.g., CCITT Group 3 or 4 compressed TIFF file) promise access by diverse applications and ensure that data will be accessible in the long term. Other applications in the organization may be developed using different tools, though they nonetheless require access to the same data. Adherence to standards allows the melting pot to function using a common denominator, rather than isolating applications and requiring special treatment of data that should be widely available across the enterprise.

WPA cannot be a technology island with limited potential for growth and restrictions on interoperability. Openness is a key to evolution, both for hardware and software, that allows the use of best-of-breed components for specialized processing areas, such as database management systems, client application development, and workflow services.

### **System Architecture and Design**

Workflow design impacts the selection of a system architecture to meet operational requirements.

#### ***Throughput/Performance/Response Time***

CPU speed often is *not* the bottleneck in imaging server system design. Although processor speed is helpful when the image application requires a great deal of database activity (e.g., lots of SQL JOIN operations), it is network expandability that most often causes a problem for image servers. If the workflow server uses a DBMS and is running on the same hardware as the image server, then a fast CPU is probably a necessity to handle all the database activity.

Consider that in an on-line processing environment, where current 10- or 16MBps networking technology is capable of supporting only a small number of workstations, the server may need to support multiple networks to reach image throughput for a large, demanding application. Often, as applications grow, the

architectural bottleneck is the server's maximum number of network interface cards and the expandability of magnetic storage, and not CPU processing power. Particularly on smaller servers, the network(s) may become saturated and the client applications unable to process work at maximum speed long before the server's processor starts bogging down with database requests.

Optical subsystem throughput is another bottleneck that is often reached before the limits of CPU or magnetic disk performance, again depending on the characteristics of the specific application. Optical drives provide an order of magnitude less throughput than magnetic disks, and the overhead of mounting optical platters and handling multiple read and write requests tends to lower overall throughput.

When considering the additional demands of workflow management (i.e., database operations), CPU performance becomes more important. Workflow managers that use database systems as a workflow engine demand more CPU performance. The ability for the workflow manager to run on a machine separate from the imaging server and other data storage software is a benefit as systems, processing volume, and performance demands grow.

### ***Scalability/Portability/Robustness***

Systems that grow over time can quickly extend beyond the capabilities of the initial hardware platform. Choosing a scalable architecture allows growth without reimplementing the application, and if the system is portable across a variety of hardware platforms, the cost of growth and change can be kept to a minimum.

The system architecture should support hardware with a range of performance that is appropriate for the variety of applications that the organization requires. Indeed, this might consist of a single hardware vendor's offering, but it is often desirable that imaging tools support different hardware and software platforms to serve different application needs, including both clients and servers.

Server platforms can span a range from single processor machines with relatively limited memory and disk storage expandability to mainframe-class servers with scalable multiprocessor architectures and vast amounts of storage. The requirements for a workgroup or pilot application will differ greatly from a larger production application, though it may be desirable to be able to use the same tools for larger applications or even to deploy the small pilot application on much larger machines. This makes it desirable that the image system architecture and resulting applications are scalable across servers with a variety of capabilities and capacities.

### ***Reliability/Availability/Security***

Other modern techniques of system reliability, data availability, and security are equally applicable to imaging and workflow applications. These include systems that support mirroring, disk striping, redundancy, and other fault-tolerant features that keep critical systems available. Imaging and workflow platforms that are based on standard components and architectures should take advantage of these system services and provide all the same benefits as "traditional" data processing systems.



## **Application Database Management Considerations**

Imaging applications require the same administrative facilities as those enjoyed by "standard" data processing applications, as they are essentially DBMS applications that manage extended data types in a similar type of database environment.

### ***Data Organization***

Imaging systems generally use relational database technology to organize and manage data. The full power of the relational database allows a wide variety of data models for an imaging application using the same inherent facilities as for non-imaging applications.

The most common data organization scheme for imaging is the folder/document/page model, but all folders are not created equal. Applications have varying requirements for managing information, and the definition of a folder can vary widely among implementations. For example, folders can contain documents and/or other folders, where documents can be single or multiple pages; or, folders can contain folders, documents, and pages simultaneously; and, folders can contain objects other than images (e.g., word processing documents, spreadsheets).

Imaging systems should provide a data model that allows for the flexible organization of data as is appropriate for the application. Although many data processing applications appropriately manage documents for each account as a folder, some applications are not best served using the folder model. For example, an application that handles ticket reservations might not benefit from using folders to manage the items, if each ticket/owner combination is referenced uniquely.

### ***Archive and Recovery***

The diversity of data types and storage media can present a considerable challenge to system backup in network imaging systems. Images on optical need to be archived along with the related alphanumeric data, sometimes requiring a sophisticated coordination effort. Sometimes the sheer volume of data causes problems with a window of opportunity to complete the archive while maintaining adequate performance; for example, if the application must run twenty-four hours a day/seven days a week.

Optically based data can be archived to either optical or magnetic media, but the longevity requirements must be considered. If a primary advantage of optical storage is durability, then the archive technique must not compromise the security of the data by relying on magnetic storage as a fail-safe backup. Incremental archives usually use magnetic tape backup, as these are short-term revolving backups. Many systems make copies of optical platters once they are full, although this can be done incrementally, as well. If images are clustered on optical platters, the application can write a sequential copy of each image on a separate platter as they come into the system.

### **Availability**

Database techniques for data availability, such as disk mirroring, striping, and other fault-tolerant features, are equally applicable to imaging systems based on DBMSs. In addition, imaging services at the application layer can take advantage of system-level services that are available to other applications running on that platform.

### **Database Reports**

Data in imaging systems are often similar to nonimage systems with the exception of rich (and large) data. Reports consisting of alphanumeric data are still required, possibly with the extended need for reports with images as well. Report writers for these systems should allow reports that mix alphanumeric and image data with the usual formatting capabilities.

---

## **14.8 IMPLICATIONS FOR OPTICAL STORAGE**

---

For all its benefits, optical storage does not offer the performance and access of magnetic disk, so image storage systems offer a number of techniques to speed up optical retrievals and overall performance. All these methods require planning in both software design and sufficient hardware resources (i.e., magnetic storage).

### **Caching**

*Optical caching* on the image server helps to speed up image insertion and retrieval to/from optical by writing images to magnetic during the operation. On insertion, workstations need not wait for images to be written to optical (particularly useful if there is a queue or the optical platter is not mounted) before continuing with the next operation. Upon retrieval, all the images in a folder can be written to magnetic disk as an intermediate step before being sent down to the workstation. This allows much faster retrieval of subsequent items in the folder, as they will be available from faster magnetic disk.

### **Prefetching**

Workstations can *prefetch* images and cache them locally to speed up the *flip rate*. For example, when a user wants to browse the first page of a document, the workstation application could automatically retrieve pages one through ten into local workstation memory on the assumption that the user will want to view successive pages of the document. When the user flips the page, another database/optical access is not required to retrieve the image, and it can be displayed almost instantaneously.



### Staging

Some applications benefit from *staging* work (images) for more efficient operation. If the processing of work items (images) is predictable, the application can automatically prefetch them from optical overnight and store them temporarily on fast magnetic disk. Note, however, that applications in which access to data is random usually cannot benefit from this technique.

Another straightforward method that can benefit application performance is to maintain images on magnetic media (possibly in addition to optical) during periods of high activity, for example, at the beginning of a processing cycle. This explicit implementation of caching allows images to be retrieved from much higher performance storage for fast access. Images need not be stored on optical during these active periods, although they could be for purposes of archival security.

---

## 14.9 ENABLING EXTANT (LEGACY) SYSTEMS

---

### Dilemma: Rest or Recreation

Imaging brings a wealth of data to the desktop, but what about the mass of data in existing systems? Image enabling through the use of imaging toolkits provides a powerful means of extending extant applications without recreating or duplicating existing data. If a major benefit of imaging is that it "brings the rest of the transaction to the desktop" (the information currently stored on pieces of paper), organizations will both want and need to apply the benefits of imaging to existing applications and systems to make maximum use of imaging technology.

In this light, open systems imaging and downsizing does not mean scrap the mainframe, but rather, deploy networked systems as an adjunct to existing host-based or other custom applications (even existing client/server applications). Rather than relocating all the corporate data that have accumulated over time in extant applications, network imaging systems can leverage that information and provide essential services for automating and improving data processing.

### Host-Based Applications: A Terminal Condition

Open systems-based imaging can be extremely valuable as a service to a host-based application, though one potentially thorny issue is the requirement to change existing mainframe application code. Mainframe code is generally complex and, for older applications, the original developers are often long gone. One powerful attribute of network imaging toolkits can be the ability to image enable host applications without changing mainframe application code by allowing imaging applications to run in a window side by side with the extant application (terminal emulation) on a workstation with a single display.

One application area that frequently benefits from image enabling is the ubiquitous mainframe-based application running on IBM 3270 terminals. While these

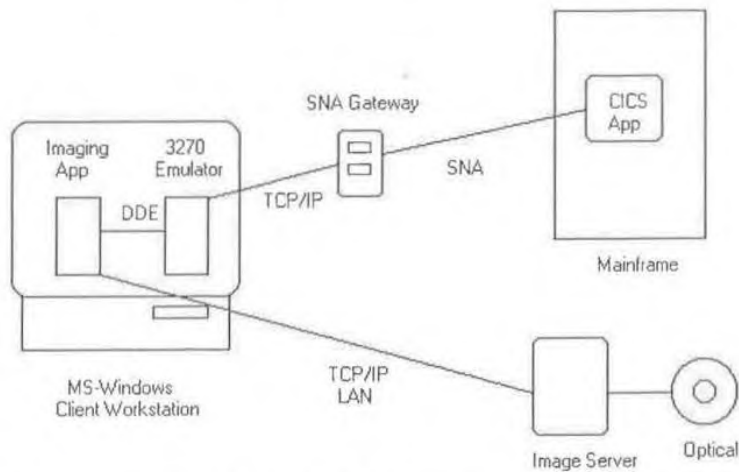


Figure 14.2 Diagram of a 3270 emulation configuration

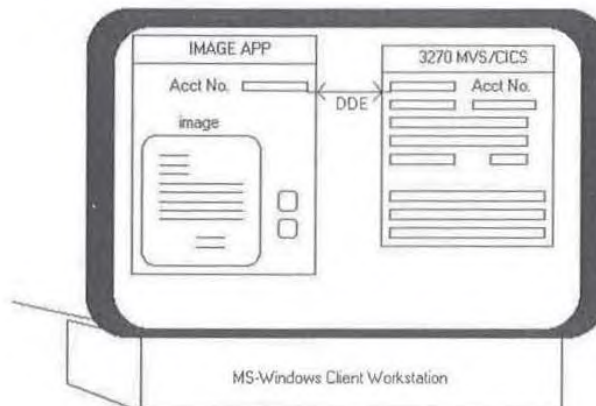
applications run in the IBM SNA world, client/server open systems imaging integration can replace 3270 terminals with client workstations that run terminal emulation software and communicate with the mainframe via communications gateways. The appropriate tools can image enable existing mainframe applications with relative ease and moderate cost (see Figure 14.2). The 3270 terminal emulation sessions on the workstation run as separate applications from the image-enabling application, and the two must exchange data to interact.

For the purposes of this example, assume that the application being image enabled is a 3270-based MVS/CICS application used for traditional on-line transaction processing (alphanumeric data) for customer service. The imaging application consists of Microsoft Windows client applications running on a local area network. The imaging application uses the TCP/IP protocol to communicate with a UNIX-based image server that allows the storage of both alphanumeric and image data, in addition to image storage on optical disk. This application uses the client/server model, where the Windows client formulates queries and requests data and services from the image server across the LAN, and the image server is responsible for processing queries, managing data, and returning results to the clients.

### Data Exchange

The imaging client application is a separate application from the 3270 terminal emulation session, and the imaging toolkit provides a means to transfer data between applications running in the client workstation. In the Microsoft Windows environment, for example, a primary facility for sharing data is DDE, which allows Windows applications to exchange information automatically.





**Figure 14.3** Diagram of a DDE screen

Microsoft Windows dynamic data exchange (DDE) is one method of transferring data and image enabling extant applications. Dynamic data exchange is a protocol defined by Microsoft through which interprocess communications are used to exchange data between applications, either synchronously or asynchronously. DDE involves a *conversation* between a DDE client application and DDE server application, meaning that a data link has been established between the two applications, and linked data will be updated automatically. The link may be either *hot* (DDE server immediately sends changed data to the DDE client) or *warm* (DDE server notifies the DDE client that the data have changed).

In the customer service application example, historical customer account data are managed by the MVS/CICS application. When a customer phones in to inquire about an account, the service representative enters the appropriate customer information into the 3270 screen and retrieves the customer's records from the CICS application. The customer account number is sent from a field in the 3270 session to the imaging application via Windows DDE, and the imaging application uses the account number as a key to retrieve the associated billing document images from the image server automatically (see Figure 14.3).

By storing both alphanumeric and image (or other binary) data, the image server may serve to enhance the existing host-based application even further. In any case, no change is made to the mainframe application to implement the imaging solution; the sharing of data occurs strictly within the client workstation environment using mechanisms provided by the imaging toolkit.

### Workflow Gateways with Heterogeneous Systems

Even though workflow routing can allow heterogeneous systems to participate in the same data processing workflow, some data translation might be needed. A workflow gateway retrieves routing data (either by reference or value) from the

workflow server on behalf of its client system, converts the data into an acceptable format, and handles any necessary batching of the data. The workflow gateway can also act as a communications gateway to allow systems that don't use the same network protocols to communicate and participate in the workflow. This requires an appropriate set of communication APIs, in addition to the workflow APIs.

### **Using CASE with Imaging Application Tools**

If imaging is really a complex form of client/server (database) application that requires a few additional technologies for handling sophisticated data types, then computer-aided software engineering principles should be applicable in the development process. Open imaging toolkits allow developers to apply various CASE (and other) methodologies when designing applications, because the same basic application development needs exist for applications that include the image data type. For example, data access, application flow control, calculation, and data presentation are functions common to both alphanumeric and image-based programs and require the same considerations for software development. Work process automation consists of a series of worksteps, which together comprise a larger metaapplication. CASE methods must be adaptable to this form of distributed application in which a large number of executables perform an aggregate work processing function.

---

## **14.10 WORKFLOW SYSTEM APPLICATION DEVELOPMENT METHODOLOGY**

---

### **Structured Methodology as Language**

The system development process is one of iterative communication. The buyer (end user) needs something built; the builder (integrator or user development team) is charged with constructing that something. The builder is more likely to deliver exactly what the buyer wants if both agree on what that something is, what it is to look like, what function it is to serve, and when it is needed. The builder will be successful if he or she and his or her staff agree on what materials are needed, the tasks and procedures for using those materials in the construction of the goal, and the methods for determining that things are going according to plan. Failure to reach informed agreement (i.e., when both sides share a common conception of what is being agreed upon) will almost certainly guarantee subsequent problems.

A sound methodology should provide terms and symbols to convey system concepts and a "grammar" (or structure) that links these into a coherent description of the system requirements and its solution. The methodology and its components should be understood by both buyer and builder and should not be susceptible to misunderstanding or misinterpretation. Such a methodology is a language, facilitating communication among all parties involved in the project.



### ***Document-Driven Method***

**Documents Record Work Accomplished and Track the Project** Documents not only record the work accomplished throughout a project's phases, but they also launch phases. The major documents generated during a complete system life cycle might include the following:

- Imaging/work process automation opportunities assessment
- Application profile
- System requirements definition (including essential model and behavioral model)
- External design specification (including user implementation model)
- Internal design specification (including processor model)
- Integration test plan and test script
- Acceptance test plan and test script
- User documentation

**Milestones to Kick-Off Phases** The system development cycle consists of a series of phases, each punctuated by one or more milestones. At each milestone, progress should be assessed, with work reviewed and accepted; plans for future phases can be reviewed, and perhaps altered, in light of the experience to date.

At each major milestone, designated team members gather to review what was built and/or accomplished during the preceding phase or subphase. For most of the milestones (e.g., requirements definition, external design), a document is reviewed; for the others (such as installation or acceptance testing), a previously reviewed and accepted document should support the assessment process directly.

### ***System Development Phases***

Every new system passes through a number of stages, from initial conception through installation and production (and beyond). An analysis and development team will not necessarily have input to all phases; in some cases (and not infrequently), management will have already decided that the solution to the business problem is an imaging/work process automation system; the major requirements for the new system may also have been defined.

Nonetheless, the development team must be prepared to address the definition and design of the new system from the outset and take it through all phases. The chronological evolution of a development project occurs in four stages or phases: feasibility, definition, design, and implementation. A fifth "phase," project management, actually spans the life of the project, yet is so distinct in its activities and deliverable products that it merits separate treatment.

**Feasibility** Prior to authorizing analysis of a new imaging system, management must determine that the work process automation and document image processing technologies have applicability in the work environment, that is, an initial determination that the new technologies are likely to solve the business problem.

*Opportunities Analysis:* The first step in such a determination is the identification of areas within the current business that are suitable for improvement by introducing imaging/work process automation. This procedure is conducted primarily via interviews with several levels of management, as well as through gathering information about the current system(s) using questionnaires and worksheets. An optional study is one that determines financial costs and benefits for each work area. Following the interviews and data analysis, the development team produces an opportunities assessment.

*Application Profile:* From the opportunities assessment, management selects one (or possibly two) application(s) of interest. The selection criteria usually take into consideration factors such as risk, visibility, payback, and window of opportunity.

The development team should prepare an application profile for each application of interest. The profile describes in very general terms the business problem(s) to be solved, briefly describes the current system (manual and/or automated), and lists the reasons why it is unsatisfactory as well as goals for the new system; finally, it describes the benefits that will be derived from the new system, amplifying the benefits list that appeared in the opportunities assessment.

The application profile will likely be a relatively short document. It is, nonetheless, essential that care and time be taken in reaching agreement with the users on each section, since this document is the first to set expectations for the resultant system.

**System Definition** During system definition, users state new system requirements by telling the analysts what the new system must do, how it must behave, what it must deliver, and how it must interact with existing systems (or those systems as they will be modified in the light of the requirements). Once the boundaries of the new system are established and the interface data transfers are defined, the development teams can model the new system as it will appear to the users.

*Essential Model:* Emphasis in this phase of development must be on securing user agreement on the functions of the new system and on how the user will interact with the system to activate these functions. The analyst teams interview the new system users to determine functional requirements (that is, how the system must respond when the user does something: add information, request the output of information, request the transformation of data, and so on). These user-initiated events usually make up the majority of the functional requirements. The system may also respond to temporal events (the passage of a specified amount of time) and conjunctions (the satisfaction of a number of prerequisites, such as the arrival into the system of all the documents required to complete a process).

Three documents comprise an essential model of the new system. An events list is the sum of all the identified events of the above three types, from which is derived



the context diagram, which graphically depicts the boundaries of the new system and its interfaces with existing systems and other external entities (e.g., other departments, customers, suppliers). The team should also document the organization of the system data to create a preliminary data model.

The behavioral model follows; it is the first view of the new system as it will behave in response to the specified events. From the events list, the analyst team creates a preliminary data-flow diagram and group system processes in a partitioned function diagram. The processes in this diagram are in turn decomposed into lower-level data-flow diagrams, whose supporting documentation includes business process specifications and a preliminary data dictionary.

*System Requirements:* The analyst team packages the essential and behavioral models with a statement of requirements for performance, training, documentation, hardware and software, backup/recovery, and data conversion, and presents these to management for review and approval as the system requirements definition.

*User Implementation Model:* Following agreement on system functional requirements, the user analysts depict the new system as it will look to the user in the user implementation model, which builds on the work already done. The process is highly iterative, and the "look and feel" of the system should be developed with a great deal of input from those who will use it. This model should contain workstation screen displays for each set of activities to be performed by the users, as well as report layouts, as appropriate. The analysts develop a set of state transition diagrams to depict the stimulus/response associations for each discrete user activity (e.g., system response when user selects an object, such as an "Update" button on the workstation screen). The database schema builds on the work done earlier in creating the data model.

*External Design Specification:* The user implementation model is packaged with statements of business constraints, technical constraints, and design constraints that serve as guidelines in implementing the desired "look and feel." The user implementation model, together with the constraints sections, forms the external design specification.

**System Design** Whereas the audience for the document deliverables issuing from the system definition phase is primarily the user community, documents developed during application design are for the most part intended for the system design and development team.

*Internal Design Specification:* There are several components of this specification, which the developers use to actually build the application specified:

**Processor model** This model partitions the processes defined in the lower-level data-flow diagrams (behavioral model) into groups according to the responsible processing activities (e.g., mainframe, data server, workstation, human).

**Control-level data-flow diagrams** These depict those user-interactive processes that have been assigned to client or user workstation processors on the processor model. They are used as the basis for high-level structure charts.

**Module-level data-flow diagrams** These are analogous to control-level data-flow diagrams, but are for noninteractive (e.g., server-based) processes. They also form the basis for the derivation of high-level structure charts for server processes.

**Module specifications** These provide a detailed narrative on what will occur in each of the processes in the control-level and module-level data-flow diagrams.

**High-level structure charts** These depict the relationships between the control and functional modules that appear on the module-level and control-level data-flow diagrams.

**System Implementation** Once the internal design specification has been approved, the development team proceeds with application coding and testing. The development method encompasses the following activities during application implementation:

- Design module structure charts
- Develop unit test plans
- Develop string test plans
- Develop integration test plan and test script
- Develop acceptance test plan and test script
- Create commented source code; test and rework
- Perform unit, integration, and acceptance tests based on the plans and scripts

**Project Management—An Ongoing “Phase”** Project management is a critical set of activities that will ensure successful execution of the above phases. Some of the project management tasks include:

- Develop a work breakdown structure
- Develop a task GANTT chart
- Assign system resources
- Create subsidiary project plans: training, documentation, installation, and so forth
- Track the project
- Develop acceptance test plan
- Install and test the system hardware
- Install the system and application software
- Perform acceptance test



## 14.11 CONCLUSION

Imaging is not a cure-all. Darwin probably said it best: "Change or die." Businesses in the 1990s are facing ever-fiercer competition to survive and thrive, and the successful ones have found that an effective way of serving external customers better is to improve how they work on the inside. Traditional business applications model paper and data flows that already existed in the organization and, as a result, business processes are often tightly integrated in the software design. And because the whole mode of operation is bound tightly to these monolithic applications, business processes often have to adapt to existing applications, not vice versa.

Converting from paper to imaged documents often boosts efficiency and cost effectiveness of an operation by as much as 50 percent simply by automating an existing process to provide better response time, lower information storage cost, and so forth. Imaging as a basic technological capability provides the *means* to route, store, and manage data more efficiently. Imaging alone, however, cannot tackle the totality of the problem of helping an organization, be it workgroup, department, or enterprise, deal with change.

Just *having* workflow isn't good enough either, especially if it's coupled tightly with the applications involved; it's too difficult to change. The key is being able to analyze the model and change the work process in a natural, interactive way, and be able to adapt the workflow to a changing business environment with minimal effort. This can happen only when an organization can separate business process workflow management from the applications involved and focus on solving the real business problems without having to recode existing applications just because the *flow* changed. In this way, work process automation allows an organization to improve its efficiency vastly more than using imaging alone.

Work process automation (WPA) is the coordination of people, tasks, and data to improve the efficiency, flexibility, and effectiveness of an organization. A combination of tools and technologies, including database management systems, graphical user interfaces, optical storage, imaging, and workflow provide the necessary pieces of the puzzle. In concert, these technologies can be used to build flexible systems that allow users and organizations to adapt.

WPA represents a fundamental change in the way an organization does business, and adaptation is an ongoing process. Companies can evaluate and reengineer the way they work, carefully and constantly reviewing the relationships and activities within the enterprise. Process change need not be revolutionary; instead, processes should evolve with minimal trauma to best serve the goals and objectives of the organization. The easier this process of process evolution is, the faster a company can address costly changes in the business environment.

WPA is *true* management information, not only about data, but about business processes, as well. MIS has traditionally provided access to and analysis of crucial business information; for example, financial data about revenue and costs, manufacturing yields, and the like. But the "data about the data" provide higher-level information about the organizational processes in place and allow companies to analyze not only *what* they are doing, but *how* they are doing it, as well.

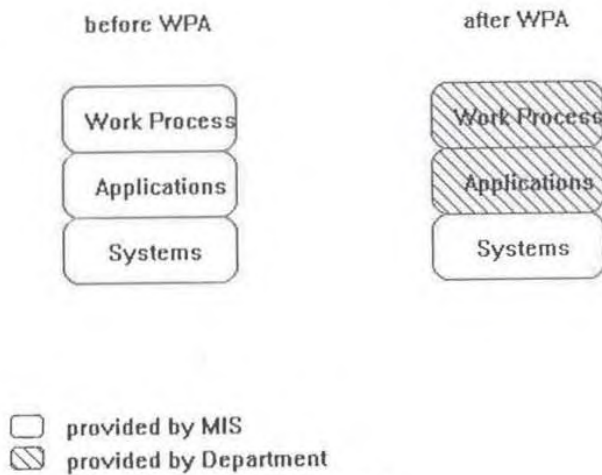


Figure 14.4 Evolution of MIS responsibilities

The role of central MIS may be very different in a work process automation environment than in a traditional centralized data processing model. Rather than providing *everything* required to implement work (data) processing systems, from application development to designing and embedding workflows, MIS may specify and/or provide the basic systems and architectures (e.g., hardware platforms, networking architecture, DBMSs) and possibly the work processing applications as well (see Figure 14.4).

This paradigm for distributed contribution will allow central MIS and department organizations to contribute most efficiently to the computing environment within the corporation. MIS might manage the basic technology of the information systems used for work process automation, and department organizations have the ability to change how they work without requiring services from MIS.

Once an organization has embarked down this path, it will find it difficult to return to the ways of the past. The freedom of decentralized computing means that organizations contribute to the system in ways that allow them to add the optimum value. Technology infrastructure and technical/integration expertise are provided by one group, either at the corporate or smaller organizational level, while the department can define the application user interfaces, work processes, and other procedural elements according to the business practices that make sense.

Imaging and workflow are effective means for sharing the wealth of data throughout an enterprise, and they allow the scope of the data in a transaction to go far beyond that of yesteryear's transaction. The mainframe may remain a requirement for much of the data processing in the world, but open systems imaging and workflow can augment its efficiency without causing upheaval—change it without touching it. This means that as application requirements grow (to imaging and beyond), organizations can add to the imaging applications rather than the



mainframe. Client/server applications can also benefit from an imaging add-on: The heart of the application is unaffected by the implementation of imaging or workflow; it is unaware of the means required to participate in this enabling technology.

The key to enabling with client/server image and workflow is the flexibility to create, change, and add to these "heterogeneous applications" that allow host-based, network, and local applications to interoperate for maximum leverage. And as the world turns away from the monolithic and toward open, synergistic, object-based environments, we will begin to see things (what we used to call applications) as providers of services to the real applications—ourselves.

---

# 15

---

## Integrated Workgroup Image-Enabled Productivity Software

---

*William A. Gelfand*

Lotus Consulting Services Group

### **Editor's Note**

*One of the rapidly emerging business support strategies made possible by the capabilities of modern workstations is in the area of workgroup computing, where files and applications are shared between individuals working on the same project, regardless of geographical or organizational boundaries. These workgroups can benefit from the integration of imaging with the other document formats that are available through the network, including word processing, spreadsheets, electronic mail, voice, and, in a not so distant future, full multimedia and video conferencing.*

*This chapter describes some of the design issues that must be considered when trying to integrate an image document system with the rest of the business applications necessary to support the workgroup. Although the discussion is limited to tools that have been fully deployed in the DOS/Windows 3.1/OS/2 environment, many of the issues carry to Windows 95, Windows NT, and UNIX-capable networks. This chapter identifies the common threads of image-enabled integrated software, understanding that, from a user point of view, deciding on the "best" workgroup productivity software may well be the result of personal considerations, such as familiarity with existing tools.*



*Consequently, from a user point of view, deciding on the "best" work-group productivity software may well be the result of two highly personal considerations:*

- *Familiarity with an existing battery of tools supporting the user's business*
- *Perceived need for "better" tools*

*If the "better" tools are from a familiar family, user acceptance will likely be higher than if they are from a different product line, requiring learning new ways of doing business.*

---

## 15.1 DEFINITION

---

Although the local area network (LAN) as we know it has been around for some time, the benefits a business derives from running its main activities on one have been inconsistent. On the one hand, if a PC application program existed to serve a particular business, more times than not, it was developed with a single user in mind. The advent of the LAN made it possible to "stretch" this program so it could be used by many people, but the program's single-user heritage always seemed to show up like an uninvited relative. And although the stretched version of the program was able to perform most tasks well, some operations, which performed flawlessly in the single-user version, turned into nightmares when attempted by multiple users on a LAN.

Eventually, programs were developed with the network in mind. These programs could be used on a single machine, but their main benefit was intended for groups of workers. Many mainframe application programs, which businesses had been struggling with for years, were laid to rest, and new programs were written and implemented using the benefits and cost-effectiveness of the PC/LAN environment.

During this time, document imaging systems, which resided on midsize or, more commonly, mainframe computers, had price tags that were almost as high as the long-term costs of maintaining them. Due to the significant processing and storage needs a document imaging system requires, these were the only platforms able to support business programs of this type. In fact, if a business could afford them, these systems did their job very well. They had relatively fast response times, excellent data storage and archive, and usually a large company to service and support them.

The PC/LAN market was speeding along, though. New and faster systems were being introduced weekly, and the storage capabilities of these computers increased dramatically. With the advent of optical storage devices that could store almost a billion bytes on one removable disk cartridge and still reside inside the PC, the reign of high-cost document imaging systems was coming to an end.

It became possible to purchase a document imaging system that would fit well within a budget and cost relatively little to maintain. Any organization could scan documents into the computer, store these images in folders and files, and index and retrieve them in a matter of seconds. The time-consuming, and hated, manual process of searching the company archives for old files was exchanged for entering information on a computer screen and watching the results pop up. For any organization needing quick and painless access to its old records, where the image can be defined as the application, this was and still is a very cost-effective solution.

---

## 15.2 THE IMAGE AS THE APPLICATION

---

For example, a law firm is preparing a case for its client. The client, a large chemical company, is the defendant in a lawsuit that claims it failed to follow the appropriate guidelines when handling hazardous chemicals. Since this type of lawsuit usually puts the defendant at risk for millions, even billions, of dollars, much effort and expense will be exhausted in order to produce, if not a victory, at least a reduction of damages. Recommendations for a document imaging system by the law firm will be taken seriously by the client even if the life expectancy of the system extends to only this one case.

In lawsuits of this type, the volume of documentation and paperwork is astounding. In one actual case, the defendant had to dust off files and documents from the early part of this century, and these were just some of the two million pieces of paper that conceivably could have been used by either party in the lawsuit. Obviously, this amount of paperwork is beyond the ability of even a team of lawyers and paralegals to deal with in a manual fashion. Even if all these documents were read by the legal team, they would find it impossible to precisely and fully cross-reference the information within these documents for the company's day in court.

In this situation, an automated document imaging filing and retrieval system makes a lot of practical and economic sense. It is practical, due to the volume of information, the need to cross-reference and index all of it, and the need to provide a fast retrieval method. It is economical, because in relation to the company's risk (potentially a billion dollars), spending half a million to two million dollars on an imaging system is minor. Even if this system is never used again, the cost can be justified. If the defense immediately fails to catch and rebut any significant statement made by the prosecution due to its inability to find contradictory information from literally roomfuls of legal documents, it could translate into many millions of dollars.

The above situation demonstrates the image as the application; that is, if it weren't for the need to automate the handling and referencing of the mountain of documentation, there would be no need for a computer system. The defendant's purchase of a computer system is not motivated by its desire for automated billing, word processing, or sending electronic mail to company employees. Its purpose is singular and will not be augmented, extended, or connected to other systems in this organization. As noted, it may never be used again.



### 15.3 THE IMAGE AS A COMPONENT OF THE APPLICATION

Imaging products that presume the image is the application generally do their job well. However, what if the image is not the application, but only a component of it? What if needs require the organization of all information related to a particular client? In one "folder" must be word processing documents, cost analysis from spreadsheets, market research from the database, and, last but not least, images of client correspondence: pictures, graphics, or any other documentation critical to this client.

What if you want to put together a subset of these components for a presentation without ever leaving the computer you are working on? What if all these components must be living documents; that is, when one component is altered, changes to the data will be reflected in the presentation finished last week? And what if various workers each have a role in producing or reviewing these components? Here, the software environment, not just the application software, as in the previous case, offers groups of people the ability to produce a presentation with the image as a single, albeit important, component of the final product.

For example, the purchasing arm of a large company requires the submittal of a number of signed documents by employees requesting a major purchase. This documentation must include financial justification (spreadsheets), usage or need justification (word processing, graphics, and/or spreadsheets), and technical details (drawings, schematics). In addition, the purchasing agents may need to route this documentation to various executives for review and signature. When all this is complete and the request is approved and purchased, tracking details, such as purchase orders, purchase invoices, shipping invoices, memos, warranties, and other documentation, need to be retained for equipment servicing or warranty reasons.

If, in this example, an integrated program was used throughout the request and purchasing process to track and record all relevant documentation, company personnel would be able at any time in the future to collect complete and consistent information on any aspect of the purchasing process. This information may be a memo from the word processor or the shipping invoice from the image processing component; the specific type of document is unimportant. The fact that all components of the purchasing process co-exist in one location and can be retrieved and analyzed in the future makes this invaluable to any organization, especially those with complex workflows.

The use of the image as only a component of the application is what integrated workgroup image-enabled productivity software (IWIPS) is all about. It is a software program or suite of programs that together allow people connected by an operational process to increase their productivity both during the process and after it when needing to research historical information.

The following sections detail how to decide if IWIPS is right for you and your business, the basic building blocks to an IWIPS system, and some add-on components to enhance and increase the efficiency of a standard system. Two products are presented as examples of the image as a component of the application to



illustrate the strengths and weaknesses of various imaging systems. And finally, a number of key issues pertinent to the success of an IWIPS are discussed.

---

## 15.4 DESIGN CONSIDERATIONS FOR AN IWIPS

---

Before any consideration is given to which document imaging system is right for your organization, these questions must be answered: "What do I want to accomplish with this system?" and "How, within my budget confines, do I accomplish this?" The first is a complex question. Besides the obvious answer—to increase the efficiency and reliability of some manual or partially manual process—consideration must be given to what gains your organization will garner, what new functionality the system will provide, how it will affect employees, and what will be relinquished.

If you have simple goals, such as organizing personal files, then the answers to these two questions will not be as rigorous for the person chartered with the responsibility of automating the organization's paper flow; saving money; and providing a safe, realistic, economical, and workable document retrieval and backup scheme.

### Define Business Operation

The first step is to map out what groups in the organization are affected by the current process. Sit down with those responsible for both the data (i.e., nonimage) portion of the business and the documents; understand the current work process: how paper and data flow into the operation, how they are manipulated, how they flow out to other groups, and how they will be used historically. If your job is to automate an existing process, this workflow detailing may produce curious inconsistencies and questions as to why the process has taken on its current form. Many times you will discover that the evolution of this process has produced some very queer beasts, which you somehow must unravel and understand.

If, due to your exceptional foresight with new technologies, you are proposing a beneficial new process, hopefully, you will have equal providence as to how it will affect both humans and machines. This new method must be both workable for those using it, economically realistic for those paying for it, and technically feasible for those installing it on the computers you currently own. More times than not, your proposal will include additional hardware and software requirements on each user's computer, in addition to at least one central site hardware configuration. Be aware that the installation of a workgroup document image processing system may change the way people use their computers even during nondocument image tasks.

In either case, it is critical after this workflow analysis that serious thought be given to improving the process and not just replacing it. More times than not, simply replacing one system with another translates into doing the same job faster. Faster is nice, but a simple translation of a manual workflow to an automated workflow



has the undesirable effect of moving warts and scars from one process to another. Just because it's being done now doesn't mean it makes sense in the future. You need to define system goals, understand the current process and all its inconsistencies, strip out as many inconsistencies as possible while retaining functionality, and finally design the system based on these facts. You are in the unique position of understanding both the current workflow and the potentials of a future automated operation. No one else has this knowledge: Use it wisely.

There are many different points to consider before investing in any document imaging system, and all of them are not detailed here; however, make sure you address at least these points listed below before investing your organization's money and operations in any product. First, understand the workflow of documents in your present operations:

- How do documents presently enter, reside, and exit your business process?
- Do these documents need to be saved? For how long?
- Who needs to access these documents? From where? When? How frequently?
- Do these documents need to be saved and grouped in a file with other types of documents?
- How much paper enters the process each day: 50, 500, 50,000 pages?
- Is the scanner you have budgeted for hardy enough for the type of documents you intend to enter into this system and the workflow you have planned for it?
- What are the bottlenecks and potential points of failure of the current system?

Second, after analyzing the workflow, determine what enhancements to the process you would like:

- Integrate these documents with a company-wide database?
- Extract the text from these documents (OCR)?
- Send and receive faxes directly into this system?
- Integrate this system with your company-wide e-mail system?
- What are the bottlenecks and potential points of failure of the new system and how can you design them out of the process?

### **Windows Workgroup Imaging System Architecture**

Connection to a network of some kind is a given these days for people who work together. This network may be local or wide area, or it may mean communicating with fellow workers through a central hub, as in mainframes and minicomputers. In either case, information is passed from one person to the next.

Limiting our discussion to the MS Windows and OS/2 environments, the choices for types of networks usually involve Novell NetWare, Banyan VINES, Microsoft LAN Manager, or NetBIOS. Any of these network choices allow for a fully

functional document imaging system. The most common requirements for Windows-based document imaging systems are detailed below. Be aware, though, that each imaging product has its own requirements, so before purchasing, it's a good idea to thoroughly check and make sure your organization's network infrastructure can support the system you have selected.

## **Client/Server Computing**

Connecting to a company network today is not as simple as it used to be. Not so long ago a standard local area network consisted of workstations that were connected via cable to a central network server that handled file storage, communications, and printing. It was true to its name; it was local. Today, the guts of the network you connect to may very well exist in a location far from where you are. Be it from a laptop computer over a standard phone line, a wide area connection over high-speed phone lines, or across the street over a dedicated fiber optic link, the local area network has outgrown its name.

From the point of view of the user, the imperceptible interaction of the document imaging system and the network, whether local or wide area, is critical to the success of the system as a whole. When the network speed is appropriate for the task, when the imaging software efficiently uses the network toolset, and when the design of the system as a whole is a result of a detailed understanding of the organization's goals and workflow needs, then, and only then, will the user feel that this is a tool that increases efficiencies and in the long run saves money. In other words, the network must not get in the way; it should be conspicuous by its absence.

The client/server model is used today both by workgroup document imaging products as well as by mainstream document imaging products, due to its inherent ability to separate processes best done on the server from those best done on the workstation. The server facilities provided by most imaging systems are: image server, fax server, OCR server, and, in some cases, a variation of a scanning server; these are in addition to the network operating system's server facilities. For example, the image server stores document images, receives requests for images, and responds to the requests by transmitting images to the requesting workstations. The workstations, on the other hand, run the programs that request, display, and print the image.

Server processes can be combined where it makes sense. For instance, the image server can be one and the same as the network file server (NFS). In such a case, the document image software manufacturer develops a special program, which is placed on the NFS to manage the storage and retrieval of images and data files for the document image software. As of this writing, Novell and Kodak are jointly developing software that will enable the NetWare network operating system to efficiently manage image files, something it can't do very well today. It is hoped that the product that emerges from this effort will provide many benefits to the NetWare user, not the least of which is a more centralized storage and backup facility for all types of files. The various servers available for a document imaging environment are detailed in the next section.



### ***The Image Server***

The most common feature of a workgroup document imaging system is its image server. It is used to separate the image file storage from the NFS, thereby increasing the flexibility of image storage management and significantly reducing the file access delay nonimage users incur when both images and data files exist in the same storage environment. An image server manages the storage of all network-based image files and works in concert with the requesting agent, the workstation. It is also specially designed to handle extremely large storage capacities and the wide variety of storage media available today. Besides magnetic media, image servers can manage WORM optical drives (write-once, read-many), RW optical drives (read/write), jukeboxes containing these media, and archive systems—most commonly tapes.

How is this done? The image server receives requests from the workstations and processes these requests by locating the desired documents and sending them to the requester. That sounds simple, and it is, except for the part about locating the documents. The main task of the image server is document management. At all times it must know the whereabouts of all documents put into its trust. In an environment where everything exists on a magnetic drive and doesn't move until it's deleted by the user, that's simple: Look up the name of the file in the internal file directory and get the file. However, when there are a number of drive types on a system (magnetic, WORM, RW optical, tape), and file locations are not stagnant (an unused file may be automatically moved to a slower medium), the image server can't just pluck a file off its disk. It must know which medium type it is on, how to access it, if that device is available, and where it is on that device. And it must have the ability to signal to both the requester and computer operator that a file was requested but its medium is not on line. Luckily for the user, this is all handled by the software itself.

Although this method of image storage requires the purchase of a separate and dedicated computer, it is currently the best solution for image file management. As mentioned, work is being done on some network platforms that will allow the NFS to handle this task, thus eliminating the need for an additional computer, and thereby reducing the price of the complete imaging system.

### ***The Fax Server***

It is quite common today to send and receive documents via fax. However, if you have ever tried to scan a paper fax, you have noticed a degradation in image quality in the resulting image document. If you then tried to OCR this document, the result was no doubt less than pleasing. Enter the fax server, a computer that manages and routes inbound and outbound faxes. Inbound, it stores the fax as an image file so that the quality of the stored image is greatly increased over the method just mentioned. As for outbound faxes, any network user can fax data or stored images or can scan and send an image directly to the fax server, usually with increased resolution over standard fax machines.

Like the image server, the fax server may be a machine dedicated to its particular task, but most products do not require this. If the volume of faxes is not significant enough, the fax hardware and software may exist on a workstation, or they may share a server with another process (e.g., fax/OCR server). It is up to the system designer to determine if, in fact, the volume of faxes sent and received versus the needs of the sister process warrants a dedicated machine.

### ***The OCR Server***

The process that converts images into text, known as optical character recognition, is important only in those document imaging systems that need to manipulate or search through the text of documents. Similar to the other server tasks mentioned, the OCR server acts as a common point of processing for all users who need its services. If your needs are simply to manage images, this option can be eliminated.

The OCR server can be set up through a hardware and software configuration or simply through software. With the advent of faster computers and cheaper memory, the trend today is toward software processing. This has an added benefit in multitasking environments such as IBM OS/2 and Microsoft Windows, since multiple OCR tasks can be spawned off either from a single request or from multiple OCR requests.

In the single request scenario, the controlling OCR software analyzes an image of text and separates it into usable pieces. Each piece is sent to the OCR engine for processing; thus, each section becomes an OCR subtask. When all subtasks are completed, the controlling process can stitch together all the pieces into one text file. This method is extremely efficient in that it uses the operating system to speed the conversion from image to text. Similarly, with multiple requests, the controlling OCR procedure can process a number of requests at once without significant delays due to backlog.

Although OCR has come a long way since its introduction, it is still the only process discussed here that cannot always be counted on. In concept, its potential is great: automatic conversion of images into computer-processable text; in reality, it is highly dependent on the quality of the image it is given. Good image quality yields high translation accuracy. Low image quality yields poor translation accuracy, which therefore results in an inability to be considered seriously for anything but ad hoc needs. In the case of fax documents, the problem is compounded by the low resolution of the fax scanning process. In time, as machines get smarter and faster, and faxes get clearer, this will be less of a problem, and the OCR process can be counted on to a greater degree.

### ***The Scanning Station***

In order to enter your documents into an imaging system, you probably will have to scan them in, and, given the proper hardware, this can be done at any workstation. However, a few questions should be asked before you bankrupt your organization



and start buying scanners for all network users: What type of documents will they be scanning? Will you need special paper handling for these documents? Will you need bulk scanning? How many people will be scanning documents? Will you need to index a folder of documents at scanning time or can it be done at a later time? There are more questions that are beyond the scope of this chapter, but the point to take home is that a good-quality scanner still accounts for a significant portion of the total imaging system cost, so it is critical to fully understand the "who" and "how" of the scanning process as it relates to your organization.

Unlike the image, fax, and OCR servers, most document image management systems (DIMS) recommend placing scanners on the workstations of those people doing the bulk of the scanning. The server concept does not exist here since a scanner is a device requiring, at minimum, some human interaction on each usage.

There does exist one product, detailed below, that comes closer than anything else to the concept of a scanning server. It allows the system administrator to configure the scanner on a dedicated computer that, when given the name of the person whose documents are to be scanned, scans the images and routes them to that person's desktop. This doesn't remove the manual step of scanning, but it comes closer than anything else to providing an economical scanner for workgroups.

### *Scanner Placement Considerations*

In an operation with a significant amount of scanning, the placement of the main scanning station relative to the storage device is important. For example, suppose a workgroup currently has its desktop computers connected to a LAN. Performance is not a big issue since small data files are the most common form of network traffic. Since they need to keep track of client correspondence, a simple workgroup DIMS is recommended and installed onto the existing LAN, complete with an image server, a workstation scanner, and software. The DIMS runs through the pilot with flying colors and the company decides to go "live" with this system. "Live" meaning that instead of scanning twenty documents a day, as they did in the pilot, 500 documents are now scanned. All of a sudden, network performance is down to a crawl. Why? Well, one reason could be that the design of this system did not take into account the effect that large image files transmitted from the scanner through the network to the image server would have on network performance. Whereas before, small data files were the main network traffic, now large beefy image files are moving through the wire using up significant network bandwidth.

There are a few simple solutions to this problem that can be defined as inexpensive and efficient. The inexpensive solution would require the scanner to be physically connected to the image server; that is, instead of the image server receiving images through the LAN, it would get them directly from the scanner. This would decrease network traffic on scanned images to zero and also speed up the storage of images on the server.

This poses other problems, though. Merging these two components would mean that either Mohammed would have to come to the mountain, or the mountain would have to go to Mohammed. Having the mountain move is fine for Old Mo', but the

person responsible for the security of the images might not like the image server out in the open. On the other hand, Old Mo' might not like being dislocated from his desk and moved to the image server's secure location. Although this is a management issue, its significance to the successful system deployment should not be taken lightly.

The efficient solution would not require any movement of people or resources, but would require that a second network "ring" connected to the main "ring" be installed. The image server, the scanning workstation, and a connection to the main network ring would be the sole components on this new "mini-LAN." All users still could access images, with the benefit that the network resources on the main ring would not be used during the transmission's intensive scanning process. Users would not even know that scanning and storage were occurring. This method does, however, have its costs in hardware, network software, and rewiring the scanning station.

### ***Workstation Client***

This is where the action is. The workstation in a client/server environment initiates almost all image storage and retrieval functions. It is the requester of image documents for display, printing, and faxing, as well as the main conduit of images that have been scanned or faxed into the system. The workstation must present a clear, easy-to-use, and consistent interface to all applications needed by the user. Therefore, for workgroup-type applications that correspond to this discussion, it is quite common for developers to use an operating environment such as the Macintosh, Windows for PCs, or X Windows for UNIX.

By definition, this discussion requires that all components of workgroup software be "integrated." Applications written specifically for each of these user-integrated interfaces must look and operate according to certain guidelines, which, when held to, provide to the user both a smooth transition from one application to the next, as well as behind-the-scenes communication among applications and data. In the case of Microsoft Windows, one might make the assumption that it is possible to use software developed for Windows in tandem with software developed for DOS. But this cannot be called fully integrated due to the lack of both a consistent user interface and a background mechanism that all applications adhere to.

## **Features for Integrated Workgroup Software**

Besides the consistent look of the software, what other features are important when deciding on the suite of integrated workgroup software?

### ***Mobility of Images and Data***

The operating environment must allow the free exchange of any type of data between different applications. For example, data and images residing in a customer file created by "Bob" can be included in a report or presentation created by



"Betty." Bob and Betty may be using completely different software, but each can use the other's data to complete his or her own task.

#### ***Ability to Display and Operate Multiple Applications Concurrently***

Since this discussion deals with software that increases a workgroup's productivity, these workgroups must use a variety of software products to complete their tasks. Many times, the user must work on a database in conjunction with image display software. It is imperative that the screen "real estate," that is, the usable area of the screen, be maximized. If not, users will be flipping between overlapping windows or resizing less frequently used windows in order to clearly view what they are working on. This has the undesirable effect of actually decreasing productivity due to the constant housekeeping on the windowed desktop.

In the PC world, maximizing display real estate means enhancing two components: the video display card and the display monitor. The standard PC comes with a VGA display card and a 14" monitor sufficient for the DOS or basic Windows applications. If, however, you want to simultaneously operate using multiple windows, then increasing both the resolution of the display card and the size of the monitor allows you to concurrently view multiple applications, each in full-screen mode. Inefficiently flipping between applications has been eliminated.

#### ***Ability to Perform DDE***

Any application used in this environment should be able to perform an operation called DDE, or dynamic data exchange. This operation allows links to be established between data files of different origin but common relation, thereby allowing multiple data files to automatically reflect a single common change.

An example of this is the linking of the data from a spreadsheet with the text of a word processor. After the initial number crunching is done in the spreadsheet, the remainder of the work most likely will be done in the word processor. If, however, changes need to be made to the numbers, it would be very convenient if (instead of having to exit the word processor, enter the spreadsheet, make the changes, save the changes, exit the spreadsheet, enter the word processor, and finally copy the changes back into the text) you could modify the numbers directly in the word processor, which would automatically update the spreadsheet with the changes.

#### ***Object Linking and Embedding (OLE)***

As recently as five years ago, a computer-based customer file could be defined as any text-based information that could be entered into the company's database. This was fine as long as the work did not involve receiving paper-based documents, producing any spreadsheets, or creating any graphics.

Today, however, what makes up a customer file includes such things as graphics, images, and spreadsheet data, in addition to the standard text-based information.

Each of these data elements calls a different place home, but they all meet in a primary location when the user wants them to. This primary location, or data hub, can be defined as an application with the ability to reference and invoke any other application so that when a secondary application appears, it shows a screen displaying the information that is directly relevant to the current task.

Let's say you work in the human resources department of a large company and maintain a database of all people who have sent you résumés. Currently, you receive a résumé and enter some of its information into your database and then put the résumé into a folder. Let's also assume that all hiring managers can access your database to view the new prospects. When a particular prospect looks good for the job, the hiring manager calls you to get the résumé. You then go to your file, and, assuming you refiled it correctly after the last request, pull it out, photocopy it, and send it off to the requester. If you put it through interoffice mail, the hiring manager will finally see it half a day to a day later.

Let's say, instead, you had a database with integrated image capabilities that all hiring managers could view. Now, when you receive a résumé, you enter a few pieces of information and then scan the résumé into the database so that it becomes just another piece of information in this applicant's file. When a hiring manager selects a prospect, he or she can view or print the résumé immediately. You are free from the paper-pushing task detailed above and no longer need to think of yourself as a clerk. This example can be applied to almost any office function where a bottleneck is created when those who need information must request it from those who have it.

## **Printing Document Images**

Now that a decent laser printer is affordable for most workgroups, it would seem that printing documents from a document imaging system would be fast and simple. Well simple it is, but fast... sorry, that's not the case. That peppy laser printer that a workgroup uses to print text-based documents will become a silent monster when asked to print images. Instead of measuring performance in pages per minute, sit back, relax, and measure minutes per page. The standard office laser printer was designed to efficiently print text, not pictures or images. This is not to say it can't do it; it can, with enough memory and patience.

For those people who always are pulling that partially printed page from the laser printer, there are products on the market that recognize this as a problem and do in fact speed up the printing process. These products require you to purchase an expansion card for your printer (only a few of the most popular printer models are supported), which essentially replaces the printer's parallel connector in the DOS and OS/2 environments. Print data are sent to the card, which formats the data, either images or text, and directs the printer to print the data, bypassing most of the printer's built-in processing. Although this may cause some irregularities with normal text printing, it is a definite improvement over the printer's internal image processing for those workgroups that need it. Images now can be printed at speeds close to the printer's maximum throughput.



Some software products require full printing engines to facilitate fast laser printing of images. Although the software manufacturer must develop its product around these components, they provide printing speeds up to the printer's advertised specifications as well as an integrated printing solution. The standard configuration includes expansion cards in both the printer and the computer to which the printer is connected, bypassing the parallel ports on both machines. One configuration even goes so far as to include the scanner in this loop, which allows immediate printing of scanned documents, making it a kind of high-tech copier. A downside to this type of software-printer integration is the imaging system's inability to use the shared print queues provided by the network software, thus delivering the above benefits only to the person whose workstation is connected to the printer.

### **Selection Criteria for Integrated Workgroup Software**

Many software manufacturers will soon be jumping aboard the integrated workgroup software bandwagon. Some might even claim to support document imaging. To be sure that their claims are accurate, ask the following questions:

1. Can you envision their product as the hub application used by all your network users?
2. Can you link this product with the specific applications you need?
3. What is the software developer's idea of integrated workgroup software? Is it fully developed and simple to use out of the box, or must it be customized through supplied or third-party development tools?
4. Other than scanners, does this product require the purchase of expensive workstation hardware? Considering the number of users on your network, is this figure within your budget?
5. Does the product support the use of a common image file by multiple users?
6. Does the product include image management and archiving features?
7. Is integrated workflow document routing important to you?

These are not all the questions you need to ask, but this list can be a general starting point for your investigation.

---

### **15.5 SOME FINAL POINTS TO CONSIDER**

---

The implementation of any successful workgroup process requires an understanding of the way it's done now. How does information flow through the workgroup? How and where is the information changed or manipulated? What

positive aspects of the current workflow will you retain, and what negative aspects will you remove, replace, or redesign? Where are the cracks and bottlenecks in the process into which both information and money vanish? What is missing or what just doesn't work in the current process from the point of view of both management as well as workers? Should the image lead the integration process, or is the image just another, albeit more complex, piece of data?

Next, it's important to understand the way it can be done with an IWIPS. How can your organization efficiently and effectively use all the new tools available in a Windows environment with existing image-enabled workgroup systems? Where can manual, often clerical steps, be eliminated and replaced with a system's innate ability to sort, process, and notify you of changing information? How can the cracks you identified be removed, and how can you guard against others from forming? How quickly can a working system be implemented? Can you use your new workgroup tool out of the box, or will you need to spend time tailoring it to your needs? And do you or your client have the time?

It is critical that management and the eventual users understand and buy into the workgroup concept. They must understand that a properly designed IWIPS is a productivity enhancing process and not just a replacement for the file cabinet. The system should not be designed just to mimic an old process but to create a new, more efficient, and simpler one. Its function and role within the workgroup should be well-defined and explained, with special emphasis placed on the added value of the image store. Too many times, those using a system with image capabilities perceive it as fun and alluring ("Wow, pictures on my screen!"), only to wake up to the realization that it was a very expensive one-night stand. The perception of this system, both within the workgroup and without, needs to be managed closely.

In many cases, the selection of an integrated workgroup software may well be dependent upon a strategic design decision as to the long-term need for maintainability of the imaging application. For instance, the choice may be between using an inherently tightly integrated environment and an integratable architecture. In the first case, the development path may be easier because of the uniformity of interface, but its flexibility may be constrained by the capabilities of the underlying engine. A good example of this approach is Lotus Notes:Document Imaging, which requires a Lotus Notes server, but has built-in image-embedding features, which make an image part of an existing Notes document. In the second case, the developer may want to use a wider range of options, but will need to assess the cost of integration. An example of a product that supports such an approach is Keyfile, which, as a Windows-based package, offers the flexibility associated with that environment. In that approach, the development of image-embedding features requires additional effort. Whether that feature is valuable to an imaging application depends upon a variety of factors, including the number of times an image must be retrieved versus the number of times textual information is sufficient.

One of the greatest benefits of integrated workgroup image-enabled products is not so much what they provide in their box, but the potential they offer through their use of the Windows and OS/2 environments. With the current ability of many



new or updated programs written for these environments to share data with other similarly configured programs, or to be the launching pad for an organization's workflow, the line between single-function programs and workgroup programs is becoming less clear. Today, the mark of a full-function product is not how much it can do, but how well it can integrate with other products. In the near future, you will not purchase a program hoping it works well with other programs, but assuming it will.

---

# 16

---

## Combining Microforms and Optical Disks

---

*Marc R. D'Alleyrand, Ph.D.*

---

### 16.1 INTRODUCTION

---

Regardless of the nature and architecture of the information network, there are instances when, for operational, legal, or other reasons, electronic imaging comes in competition with conventional and computer output microform-based systems.

Because of the delays associated with microform production, most microform usage is found in either look-up or archival applications. In contrast, optical disk systems provide immediacy of access, making it possible to extend the use of imaging to transaction and other on-line applications.

There are two different options for capturing and retrieving documents when microforms are involved, and each has its advantages over the other. Everything else being equal, when requirements are for immediacy and speedy communication, there is a strong argument in favor of optical disk installations; when archiving is a major consideration, a good case can be made for the use of microforms.

In many cases, however, an analysis of information retrieval requirements throughout the life cycle of a document often shows that both are desirable to support the business objectives of an organization. For instance:

- In financial applications, optical disks allow the design of integrated information systems, where data extracted from digitized paper records are processed in a cost-effective manner. The dissemination of reports generated as a result of this processing may be more cost-effective on microform than if retrieved through network imaging stations.
- In legal areas, optical disks allow efficient information retrieval. However, the cost of storing inactive files may favor the use of microforms for long-term retention.



- In sales and marketing activities, optical disk systems can provide quick access to information in support of the organization. But, because the need for instant retrieval begins to fade, microforms provide an inexpensive storage medium for archival data.

Clearly, optical disks and microforms can often be seen either as complementary or competitive rather than exclusive media (see Figure 16.1), and many organizations actually use both. There are essentially three ways to generate and use images from documents. Figure 16.2 shows that documents, once captured, can either be printed on paper, displayed on a computer screen, or printed on film. Operationally, these different methods are not equivalent.

The difference in information usage reflects different organizational responsibilities; therefore, it is not unusual that documents are captured independently on both optical disks and microform. This results in a substantial duplication of effort in preparation, indexing, and quality control. A better solution is to replace this dual image capture procedure by a process that combines recording in one medium, either on film or on optical disk, followed by an automatic conversion to disk or to microform. An alternative method, concurrent image capture, uses specialized equipment that contains both a scanner and a microfilm camera. Whatever the approach, the result is the creation of interdependent microform/digital image sets (MDIS). This chapter compares these three methods from both an operational and financial perspective.

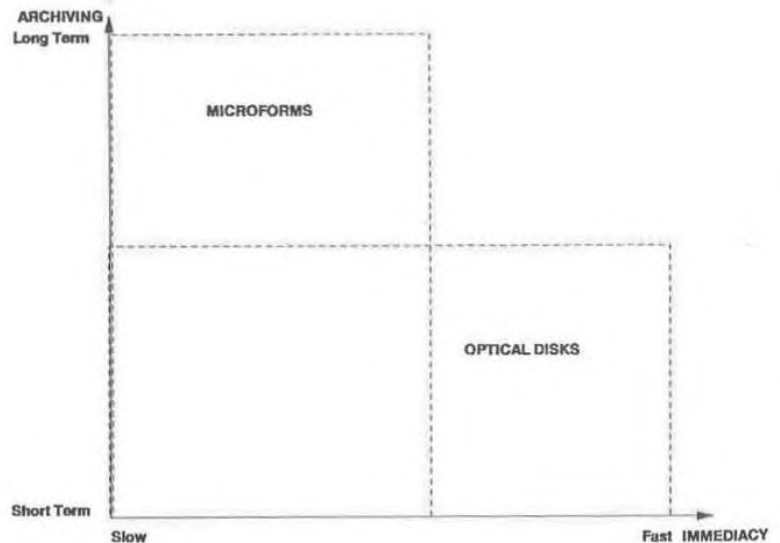


Figure 16.1 Comparison between application areas for optical disk and microforms

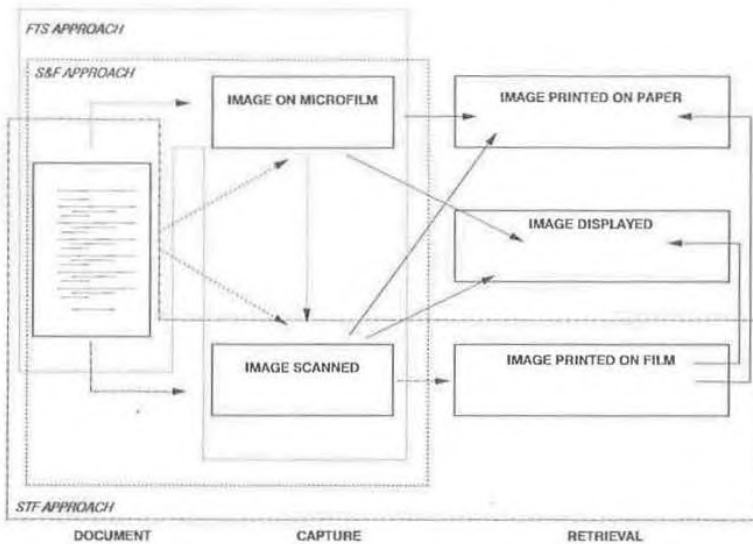


Figure 16.2 Paths for document retrieval

## 16.2 GENERATING MDIS

Three MDI generation methods have developed to meet diverging business requirements. Organizations that have an initial inventory of microforms and want to take advantage of the operational benefits of digitized images may choose to use a "film-then-scan" (FTS) approach. Operations that require a chronological audit trail and will generate simultaneous recordings on disk and microforms probably will use a "scan-and-film" (S&F) technique. Finally, organizations that need accurate backup files on microforms may want to use a "scan-then-film" (STF) methodology (see Figure 16.3).

### Traditional Method: Film-Then-Scan

The film-then-scan method is often found in organizations that have geographical or institutional obstacles that separate direct interaction between microfilming and optical disk areas. This method is also used in situations where there is a substantial inventory of filed material that needs to be converted to a digital format. As such, its implementation does not require changes in microfilm or scanning procedures, because these two processes can proceed independently. Manufacturers of film scanners for such a method include Meckel and Sunrise for roll film and fiche and TDC for fiche formats.



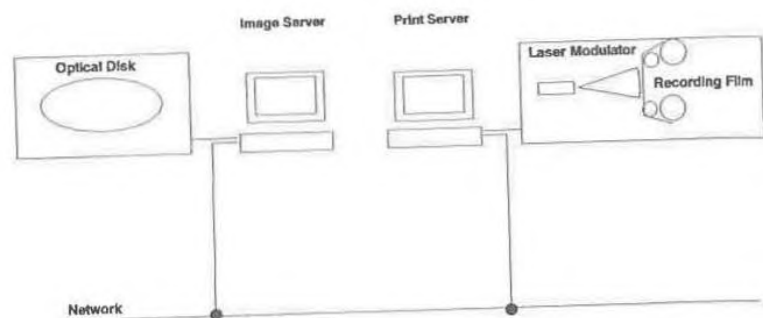


Figure 16.3 Principle of a film recorder

### Approach

In the film-then-scan approach, microforms that have been generated elsewhere are processed by a microform scanner (see Figure 16.4). Standard workflow can be described as follows: Documents are recorded on a conventional rotary or planetary camera. The resulting film is processed and inspected, and any filming errors are corrected in retakes. Once the film has been accepted, it is sometimes duplicated for safety and then sent for scanning. The scanned images finally are submitted to quality control and indexing procedures to ensure retrievability of the individual images.

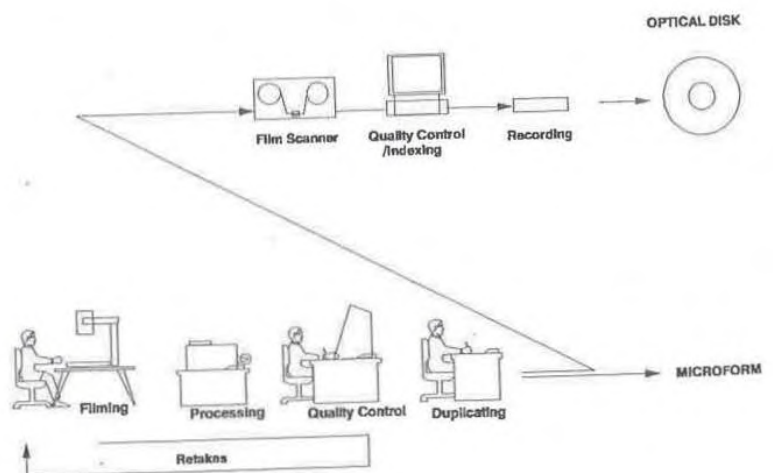


Figure 16.4 Principle of film-then-scan approach

### **Advantages**

Advantages of the film-then-scan method are: It does not disrupt existing operations, and it provides an efficient method to digitize manual files. Assuming that the contents of the film meet accepted quality standards, the conversion of the files to a digitized format can be done in an unattended mode, which saves on operator's expense. In addition, if the microforms are part of a computer-aided retrieval (CAR) system, the conversion may be facilitated by importing the index already used for the CAR system into the digitized environment. Even if the film quality is below standard, most film scanners can be equipped with an image enhancer so that marginal images, such as records of multipart forms, can be brought back to a level where they can be used.

### **Drawbacks**

The major drawback of the film-then-scan method is the same as that associated with microfilming in general—namely, the inability to detect filming errors as they occur. These errors include missed pages or overlapping images; poor document registration; and poor-quality originals, such as receipts printed with dried-out ribbons and multipart forms. Such errors are often objectionable, and it may be necessary to refile the rejected material at significant additional expense. The total unit cost of such a retake, including searching for the originals and inserting the corrected images in the filmed record, either by splicing or through indexing, may be an order of magnitude larger than the cost of image capture alone. In addition, indexing is often a dual operation, as microform indexes seldom meet the precision required for the retrieval of images on optical disks, thus requiring a second indexing activity at time of scanning.

### **Concurrent Method: Scan-and-Film**

While film-then-scan methods can be implemented in separate sites, scan-and-film techniques require the use of dual-purpose equipment, containing a scanner and a microfilming camera. These devices, most of which are of recent design, are manufactured by a variety of vendors, including Kodak, Bell & Howell, and Photomatrix for roll film, and TDC and Staude for microfiche.

### **Approach**

In the scan-and-film method, the image of a document is captured on film and scanned at the same time. Figure 16.5 outlines two different uses for this approach. Because there is no control on the quality of the microform that is produced during scanning, there may be rare conditions where the microform copy must be of acceptable quality, requiring a separate optional quality control and retake. Depending on the design of the document system, quality control of the recording may



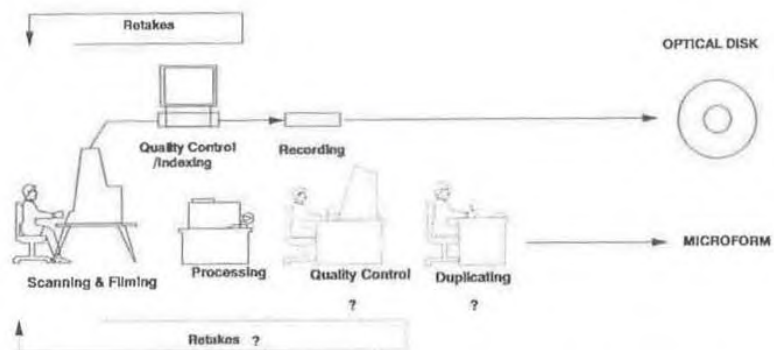


Figure 16.5 Diagram of concurrent scanning and filming approach

be done on the fly by looking at a monitor attached to the scanner or at a later time. Because of the linkage between the scanner and the microfilm camera, any errors requiring rescanning result in a duplicate image on film.

#### *Advantages*

The major advantage of this technique is that it is possible to identify document capture errors, such as skew and other misfeeds, as they occur. This allows immediate correction of the scanned image, and, depending on the economics of the correction process, may reduce the volume of refilming that would be required under the film-then-scan approach. Furthermore, it is possible under certain circumstances to calibrate the scanner so that it emulates the microfilm camera within a given range of exposure. Finally, because filming is a sequential operation, the resulting microform can be used as an audit trail.

#### *Drawbacks*

Because each scanned image, acceptable or not, generates a microfilm record on a scan-and-film device, it is not possible to prevent unacceptable images from being recorded on film. The significance of this depends on the purpose for the film copy. For instance, when the film is created only for audit trail purposes, duplicate frames recorded as the result of retakes are of little consequence. However, if the film is to be used as the official copy of record, there may be a need to cross-reference images on disk with their microfilm duplicate frames. Under these conditions, it may be necessary to develop and reconcile two different indexes. Last but not least, scan-and-film devices are in the high end of scanning equipment prices. Therefore, the use of multiple scan-and-film devices may prove to be more expensive than other imaging equipment.

### Advanced Method: Scan-Then-Film

Scan-then-film (STF) methods convert digitized images to film without the drawbacks associated with conventional microfilming. They use laser-driven devices called microform printers or film recorders. These devices are produced by manufacturers specializing in computer output equipment, such as iBase and Anacomp for black-and-white imaging. For color output, it is possible to use more expensive equipment by Dicomed and III, which use a cathode ray tube (CRT) instead of a laser source, at a significant reduction in recording speed.

#### Approach

Scan-then-film methods, depicted in Figure 16.6, capture digitized images of documents on an appropriate medium, such as optical disk or DAT, and use that medium as input to a printer specifically designed to record digital signals onto film.

At the present time, there are three types of recorders on the market, using lasers, CRTs, and electron beam recording (EBR). Most commercial recorders for document systems are laser-based machines. CRT and EBR systems, which are much more expensive to acquire and run, are used for specialized applications, such as high-end monochrome or color publishing. Because current laser-based installations operate in a binary mode, they are not well adapted to the reproduction of continuous-tone materials. CRT- and EBR-based machines may be appropriate for that purpose, but their speed may be too slow for a given application, and their operating costs are substantial.

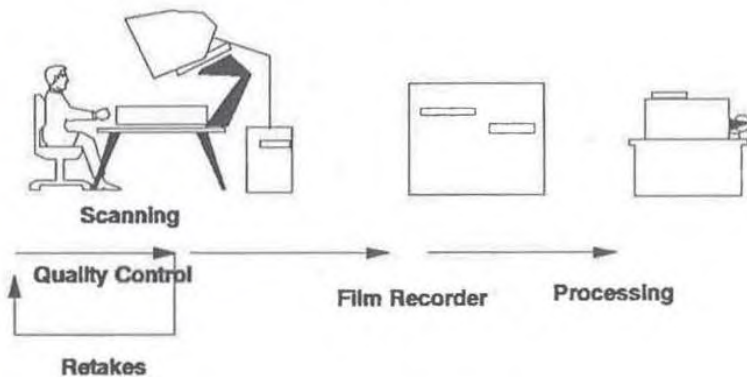


Figure 16.6 Diagram of a scan-then-film operation



### **Advantages**

The advantages of scan-then-film methods are derived from the superior quality of digitized images, which are typically electronically enhanced by the document scanner before being recorded on the digital medium. In addition, because the disk contains only valid images, the copy to film will be comprised only of relevant frames. This reduces the need for film quality control mentioned for the scan-and-film method. Furthermore, because the digitized images are fully indexed, there is no additional work required to generate an index to scan-then-film films. Finally, because the transfer is done under computer control, it is possible to merge data from different sources on the same film, including sequential numbers, electronic logos, title frames, and other machine-readable data. This provides the flexibility necessary for the production of composite documents for various applications such as desktop and workgroup publishing and reports on microform for distribution purposes.

### **Drawbacks**

Film recorder resolution is determined by the number of addressable spots, thus limiting the maximum resolution for a given image size. At the present time, most film recorders use bimodal monochrome lasers, which, because they are either on or off, require special dithering drivers to emulate the continuous-tone capabilities of conventional microfilm. Speed is also an issue when using CRT-based devices.

---

## **16.3 COMPARISON BETWEEN THE THREE METHODS**

---

Depending on the situation, there may be overriding organizational considerations that dictate one or the other method. For instance, institutions such as libraries, which already have extensive microform collections that they make available to the public, would use the film-then-scan technique; an operation with a high volume of transactions would benefit from a scan-and-film approach; and a micropublishing house might need a scan-then-film installation. However, everything else being equal, the three methods present different technical attributes. One of the major elements of comparison between the three methods is in the area of image sharpness as it is presented to the user.

The industry uses different measures to quantify image sharpness. This reflects the fact that while maximum original size and optical quality of the image capture device are determined by the image capture system, the size of the displayed or printed document depends on the display or printing device. Therefore, an objective comparison between the three methods requires a precise qualification of the output image. This issue is discussed below in more detail.

## Quality

Due to the difference in recording processes, the quality of digitized and nondigitized images varies with the image capture technique. For instance, film-then-scan methods usually provide lower-quality digital images than scan-then-film because the digital images are generated from film (i.e., is an additional generation) rather than from the originals. At the same time, the quality of images generated through scan-and-film may not be as good as either, because the microfilm camera in a scan-and-film device is usually of a rotary type, which yields a lower resolution.

The comparison criteria for the three methods focus on the quality of both the digitized image and the microform produced. This includes image sharpness, contrast, and integrity of the imaging record.

## Image Sharpness and Resolution

The method for measuring image sharpness depends on the medium and recording process. For microforms recorded with a conventional camera, image sharpness is usually expressed in terms of line pairs per millimeter (lp/mm) and refers to the number of standard black-and-white lines that can be comfortably distinguished with an appropriate microscope. For scanned images, it is expressed in terms of dots per inch (dpi) and describes the number of picture elements per unit length that can be differentiated on the original or the number of separately addressable picture elements that can be printed or displayed per unit length. These two methods of measurement are diagrammed in Figure 16.7.

A good resolution value for office documents on microforms is 140 lp/mm for images captured at a linear reduction ratio of 24 (noted 24X). This corresponds

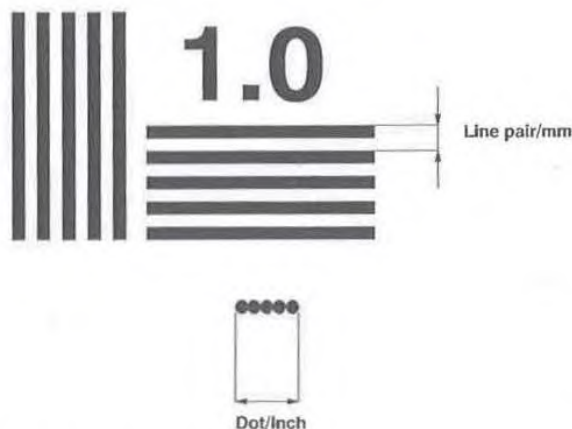


Figure 16.7 Defining image sharpness



numerically to an approximate scanning density of 300 dpi for an 8.5" × 11" document. However, differences in image characteristics require a scanning density of 400 dpi to match the resolution of a 140 lp/mm.

With film-then-scan and scan-and-film techniques, sharpness is determined by the quality of the microfilming camera. Two types of cameras are used most often to generate the film. Rotary cameras, generally less expensive to operate, also inherently have a lower resolution. Planetary cameras yield film at a higher-resolution recording, but at a higher cost than rotary filming. A third type of camera, called step-and-repeat, uses planetary film transports.

For applications that require image processing, such as OCR, only film generated on planetary cameras should be considered, and its resolution should be at least 140 lp/mm at 24X. However, one should be aware of the economics of OCR techniques; correcting recognition errors may be more expensive than direct key entry, and this may eliminate the film medium altogether. Even if the type of document is such that film processing can be cost-effective, scan-and-film solutions may not be acceptable, due to the fact that currently, scan-and-film devices on the market use rotary cameras, yielding film of lower resolution than if generated by a planetary camera or with a high-resolution film recorder.

For microforms generated by the scan-then-film method, resolution also depends on the recording process. CRT-based recorders are limited by the imaging screen definition. Resolution of laser recorders is limited by the size of the recording spot. A quick calculation shows that a recording at a reduction ratio of 24X, yielding an image sharpness equivalent to a 300 dpi scanning density, requires a spot of 3.5 micron, which is well within the capabilities of commercially available laser recordings. The same image recorded at a 96X reduction would require submicron spot size.

### ***Contrast***

The ability to record different shades, one of the strong advantages of microforms, is not a primary consideration for most present digital applications. Without special treatment, gray scale reproduction is potentially better on microforms generated by film-then-scan and scan-and-film techniques than by scan-then-film methods. For scan-then-film applications, CRT-based machines are better adapted because electronic imaging is essentially a binary recording. The use of dithering techniques to simulate shades of gray on a laser recorder, as well as the possibility of modulating a laser beam, should give laser-generated scan-then-film film similar tone rendition capabilities as film-then-scan or scan-and-film.

### ***Recording Integrity***

By necessity, correction of conventional microfilming errors requires refilming of correction frames and their insertion (splicing) in the appropriate place on the microform record. Error correction on microfiche requires the refilming of entire cards. The resulting loss of physical integrity of the film has significant importance when film is used for archival purposes, because the strength of most splices decays

over time, yielding disconnected strips of film. The presence of splices may also raise issues of authenticity of the corrected frames and weakens the argument for using microforms as archival evidence. Recording integrity can be preserved in the scan-and-film technique if the film is used only as an audit trail, and not as an active retrieval tool, thus eliminating the need for splices.

In contrast, scan-then-film methods provide the ability to generate splice-free microfilm, because film recording is performed only after the image capture process has been completed and verified.

## **Operational Considerations**

In many applications requiring the production of microforms, quality criteria are overshadowed by operational issues related to workflow and efficiency. The following discusses some of the major factors that influence the selection of one method over the others.

### ***Indexing***

Retrieval of digitized documents requires a precise directory of the recorded electronic document; therefore, indexing designed for microform applications is often not detailed enough for digitized images. Consequently, if documents have to be stored both on film and digital media, indexing first for the digital medium and then extracting the data required for the microform application will eliminate the need for two different indexing activities.

For that reason, scan-then-film methods have a cost advantage over the other techniques. The automatic extraction of microform index data from the digital image base can be performed automatically, so the creation of that index is practically "free," requiring only that the extraction process replace the address of digital images with the location of the corresponding microform. This replacement is, by nature, automatic with the scan-then-film method. This is not the case in a scan-and-film system, because occasional insertion of correction frames in microform distorts the one-on-one relationship between the digitized image and film. This distortion must be manually identified and corrected in the index, causing additional costs to ensure accurate retrieval.

A second cost benefit of the scan-then-film and of the electronic portion of the scan-and-film technique is that it allows "heads-up" indexing, where the process is done from the digital image as it is displayed on a computer screen. This process has been reported to be 30 percent faster than the more traditional "heads-down" indexing from original paper documents. In the case of a film-then-scan approach, this benefit cannot, obviously, offset the cost of initial microform indexing.

### ***Quality Control***

Quality control is probably the second largest expense in microfilm production. As such, the scan-then-film method is the most effective, as it ensures that the recorded image is of an acceptable quality before it is committed to film. The immediacy of



the image display facilitates a concurrent quality control, which, in the case of film-then-scan methods, is not possible.

Such concurrent quality control is possible to some extent with scan-and-film systems, and some errors can be caught immediately. However, practice shows that some errors will be caught only during indexing, which usually does not occur during the scanning process. Such errors require the creation of new frames that will be physically located on a different piece of film.

### ***Titling***

Microform standards require that they be identified by eye-readable title frames. In the film-then-scan approach, these are created by printing large-size targets, which are interfiled manually with the documents being filmed. In the scan-and-film and scan-then-film methods, these targets can be generated directly on the computer that is driving the scanning device, thereby reducing the production cost of the digitized record. However, unless the scan-and-film device has electronic titling capability, these targets still need to be inserted in the workflow.

### ***Format***

Microforms are used in different formats, including standard roll film, microfiche, and aperture cards. Depending on the application, one format will be preferred to another. For instance, office applications use mostly 16mm roll film; micropublishing mostly microfiche; and engineering drawings either aperture cards, microfiche, or 35mm roll microfilm.

In the film-then-scan approach, the creation of the microform requires a different type of camera, and the resulting microform can usually be fed to the same microform scanner by simple mechanical change of film transport. In the scan-and-film case, recording devices are, by construction, single-format machines, for either roll film, microfiche, or aperture cards.

In contrast, film recorders used for scan-then-film methods are multiformat, allowing the generation of film, microfiche, and aperture cards by simple substitution of film transport.

### ***Management***

The management of a document capture process requires keeping track of the status of the various processes involved, from document preparation to duplication and distribution. While the film-then-scan method can be automated as well as the other two, the scan-then-film method is the most effective of all, because the film is created only when the document capture cycle, including indexing and quality control, has been completed, thus simplifying management of the overall workflow.

In addition, because all data in a scan-then-film installation are kept in a digital form until final output, they can be routed electronically under workflow control. This makes it possible to reach higher throughput than with the two other methods,

corresponding to lower personnel costs for delivered images (i.e., initial capture and retake).

Finally, from a production backup standpoint, it is much easier to back up a film-then-scan or scan-then-film operation than a scan-and-film installation. In the first two cases, failure affecting document capture can be backed up at a relatively low cost by standby units, and failures of microform scanners or film recorders can usually be repaired before backlog from document capture becomes objectionable. In the scan-and-film method, deficiency in either type of recording results in the disruption of the total workflow, unless expensive standby equipment is available.

---

## 16.4 COST COMPARISON BETWEEN THE VARIOUS METHODS

---

### Equipment

From an equipment standpoint, both film-then-scan and scan-then-film techniques can operate in a cluster mode, using groups of input devices (microform cameras and scanners) to feed their output into one conversion device (microform scanner or film recorder). These techniques can therefore be adapted to large production installations with centralized or decentralized scanning. In contrast, scan-and-film methods that use camera-scanning assemblies operate best in a centralized scanning environment.

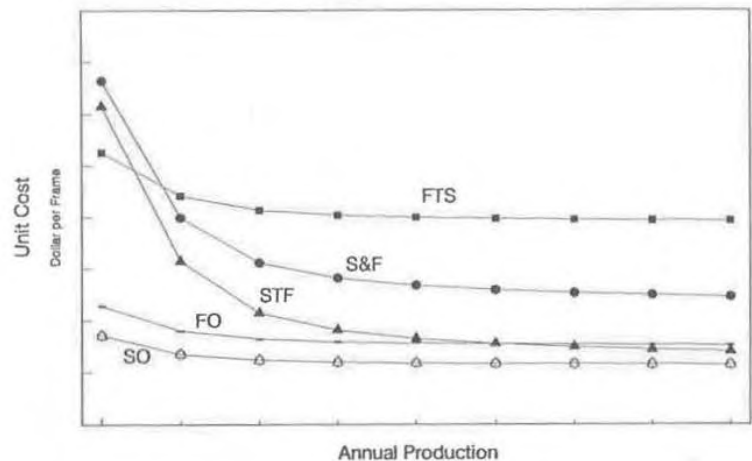
Consequently, the share of the conversion device (microform scanner or film recorder) in the overall cost for scan-then-film and film-then-scan techniques will likely be lower than for scan-and-film methods as soon as production levels require more than one scanning workstation.

### Personnel Costs

From a staffing standpoint, all three methods require document preparation, the cost of which is approximately the same for all. Indexing is more expensive for film-then-scan methods than for the other two, due to the need for a dual index for both the film and later for the digitized image.

Cost comparison between the three methods involves quantification of the different operations in the entire workflow, from document preparation to final delivery of film, including chemical processing, quality control, and equipment depreciation. This quantification is strongly application-dependent. While its precise determination may require a thorough cost analysis by a consultant familiar with both microfilm and electronic imaging workflows, it can be shown that in most instances, there is a level of production above which lower personnel costs resulting from a higher throughput of scan-then-film equipment more than offset the higher equipment costs, thus making that method the most cost-effective of the three. This is demonstrated in Figure 16.8, where, using the production figures corresponding to labor-intensive document capture and conservative microfilm





**Figure 16.8** Cost comparison between microform and digitized production methods

equipment costs shown in Figure 16.9, the cost per frame for scan-then-film, initially higher than all other techniques except for the scan-and-film, eventually drops below the cost of conventional microfilming (FO) due to the higher staff productivity. To provide a balanced comparison, Figure 16.8 also provides a comparison with scan-and-film, film-then-scan, and scan only (SO).

## 16.5 OTHER CONSIDERATIONS

Regardless of cost advantage, the identification of the best method depends on application requirements, which will need to be weighed. A method for selecting the best technique may be to weigh the various characteristics for a particular operation, using a ranking similar to the one provided in Table 16.1.

Here again, because the choice of selection criteria and their relative weight must reflect business priorities, it may be necessary to secure the help of an experienced consultant. Some of the most common selection criteria are given in the next section.

### Equipment Sharing

The model used in Figure 16.9 assumes a one-on-one relationship between film and scanners, which may overestimate the equipment costs of film-then-scan and scan-then-film methods for higher production volumes. Often, an issue will be raised as to the value of shared equipment.

WORKFLOW ASSUMPTIONS		
Preparation		
Unstapling	150	Documents/hr
Marking	100	Documents/hr
Batching	100	Documents/hr
Indexing	150	Documents/hr
Preparation Duplicates	25,000	Images/hr
Filming	500	Pages/hr
Scanning	2,000	Images/hr
Processing	10,000	Images/hr
Duplicating	25,000	Images/hr
Quality Control Film	1,500	Images/hr
Quality Control Duplicate	30,000	Images/hr
Quality Control Scan/Indexing	100	Documents/hr
Retake Film	100	Documents/hr
Recording	2,500	Images/hr
Packaging/Shipping	25,000	Images/hr
Hourly Rate (with fringes)	\$18.00	
Document Assumptions	5	Pages/document
	5%	Error rate
Equipment Costs	Capital Cost (\$)	
Microfilm Camera	10,000	
Film Processor	10,000	
Camera/Scanner	100,000	
Scanner	15,000	
Film Recorder	1,250,000	

Figure 16.9 Hypothetical production assumption for cost comparison

Only scan-then-film techniques can offer the benefits of true network document capture installations, as scan-and-film equipment cannot share its microform portion. In the case of scan-then-film installations, it can be shown that they can yield lower overall cost per image than the use of stand-alone techniques. Furthermore, network installations make it possible to rearrange workflow to accommodate operational requirements, especially in case of equipment breakdown.

### Database Reconciliation

Another cost element that further increases the cost advantage of the scan-then-film technique is the cost of matching microform and digital databases. This reconcili-



**Table 16.1** Summary of features of microform and digitized image production methods

<i>Features</i>	<i>Rank</i>		
	<i>FTS</i>	<i>S&amp;F</i>	<i>STF</i>
IMAGE QUALITY			
Resolution	1	3	2
Contrast	2	3	1
Recording Integrity	2	3	1
OPERATIONAL CONSIDERATIONS			
Indexing	3	2	1
Quality Control	3	2	1
Electronic Titling	3	2	1
Formats	2	3	1
Management	3	2	1
Costs—Low Volume	1	2	3
Costs—High Volume	2	3	1
OTHER CONSIDERATIONS			
Data Extraction	3	2	1
Database Reconciliation	2	3	1
Records Management and Archiving	3	2	1
Document Composition	3	2	1

ation cost is nonexistent in the scan-then-film method because the film and its index are generated from the digital database itself. Reconciliation is an added cost for the two other methods, and, for large databases, that cost can be quite substantial.

### Records and Archive Management

Because microforms generated in a scan-then-film mode can be produced to represent the most recent version of a document file, that technique is appropriate for file consolidation for archiving, using the digital database index to eliminate automatically unnecessary documents and generate archival records on film and produce the appropriate index at no additional cost.

In addition, film recorders produce high-resolution images; therefore, the archived files can be scanned back into a digitized system and produce images with acceptable quality. Furthermore, if the index has been preserved in a machine-readable form, this rescanning can be done with minimal intervention.

### Data Extraction

In many applications, data contained in captured images must be extracted, usually by optical character recognition, for future processing. Unless the microform has been generated specifically with that requirement in mind, image quality often will not be sufficient to ensure acceptable levels of accuracy. Therefore, only the digital recording of a scan-and-film device and scan-then-film installations will be acceptable. As already mentioned, economic consideration of the relative costs of OCR error correction and keyboarding may be a determining factor for this selection criterion.

### Document Composition

In applications involving heavy document editing and final output to film, scan-then-film equipment offers a unique advantage over the other two. In publishing in particular, it is often necessary to merge documents from different sources and reformat them. The resulting document, such as an updated technical manual, is then reproduced for mass distribution.

Because scan-then-film is mostly an electronic process, it allows the use of desktop and workgroup publishing techniques to create final documents, which can then be output in an appropriate format for micropublishing or photocomposition.

### Selection of Best Method

From this discussion, it is apparent that the selection of an appropriate method for creating microform/digitized image sets is a complex process, in which operational and other benefits must be carefully weighed. In particular, one must be aware that the potential cost advantage of one method must be fully assessed over an entire product cycle, not just during the production of the set. For instance, in publishing applications, the flexibility offered by electronic editing far outweighs, beyond a certain production level, the higher initial equipment cost of scan-then-film systems. Using the same data as those used for the model depicted in Figure 16.9, it can be shown that scan-then-film techniques can become competitive for production levels approaching 50,000 pages per month when operational criteria, such as quality, flexibility, and management of the workflow, are added to pure cost-per-frame considerations.

---

## 16.6 APPLICATION TO COLD TECHNOLOGIES

---

Computer-output-to-laser-disk (COLD) technologies are the computer data equivalent to source microfilm, intended for the storage and retrieval of computer



reports. In COLD systems, computer reports that traditionally would have been output to microform through computer output microform (COM) technologies are directly recorded on optical disk. As such, it is vastly superior to earlier inefficient and low-quality digitizing methods in which computer printouts were scanned so that their electronic images could be searchable.

When compared to computer output microform, COLD technologies benefit from the same medium advantage that optical disk has over source microfilm. Cost justification of a COLD installation, based on a comparison between cost-per-output pages, is a straightforward operation.

Currently, there are several COLD packages on the market and, depending on their features, they can provide intelligent information retrieval. In fact, COLD software are database management systems bundled with optical disk drivers designed to read input computer data submitted in a report format and to display them in a variety of ways, including traditional tabular layouts. As such, they can be found in a multiplicity of configurations, for either direct computer input or for batch processing from magnetic tape, the traditional input medium for COM recording.

There is, however, in spite of its technological similarities, a major difference between COLD and document imaging. In contrast with scanned images, COLD data are received and stored in a machine-readable form. Any graphical attributes, such as logos and forms, are added through external merge. Because of this difference, file formats between COLD and scanned documents are different, and the dissimilarity in retrieval patterns makes it rarely reasonable to store both types of data together.

---

## 16.7 CONCLUSION

---

As it has been shown, there is no fundamental reason why microforms may not be part of a full imaging system. As a backup, they offer the possibility of low-cost emergency support. However, it must be understood that such a backup mechanism must be designed with consideration for recovery. Under these conditions, the design of the imaging system must take into account the processes that will be necessary to "reload" the central system, once the emergency has passed.

---

# 17

---

## Deploying an Imaging Network

---

Mark Roy

Bitex Corporation

### Editor's Note

*Imaging is now a fait accompli, and its implementation in an organization network is essentially an integration effort with the rest of the organization's communication system. Two considerations, internetworking and inter-operability, are at the heart of such an integration.*

*Internetworking relates to the combination of local area networks and wide area networks. Interoperability refers to the ability of users of an internetwork to exchange data in a transparent manner, without consideration of intermediary hardware and network systems and protocols. At the present time, the only reliable methodology for designing interoperable systems is to use the ISO seven-layer model, described in Chapter 3.*

*Most important for interoperability in local area networks are the services offered by layer 3 (network) and layer 4 (transport) protocols. These services will often be bundled with software that control higher levels, such as Xerox Network Service (XNS), which roughly spans between layers 3, 4, 5, and 7, or TCP/IP, which covers all layers from 3 to 7.*

*Ensuring interoperability with mainframe-based networks requires special handling, depending on their architectures, as explained in Chapter 9. Successful interoperability depends on two different considerations:*

- 1. Ability of two different protocols to coexist on the same network.*
- 2. Ability of two different implementations of the same protocol to interact.*



*There is no general methodology to ensure such interoperability besides a systematic compliance testing, using all possible combinations of software available on the networks. Such testing will be greatly facilitated by (in fact, will often require) the use of powerful network and protocol analysis tools.*

*This chapter addresses some of the issues that must be considered when combining LANs and WANs, using the example of an insurance claims processing application to do so. Its perspective can be used for many enterprise-wide implementations.*

---

## 17.1 GENERAL SYSTEMS DEPLOYMENT CONSIDERATIONS

---

Among the myriad considerations necessary for the successful deployment of an imaging system is the impact that imaging will have on existing platforms. Regardless that LAN technologies may currently exist in an environment, it's likely that the increased demands of imaging will require major changes in platforms and the infrastructure. The technologic criteria notwithstanding, close attention to the details of personnel ergonomics and the process enhanced or replaced by imaging must be analyzed. In many cases, as stated throughout this book, imaging technology dramatically changes the means by which a company conducts its business. In the discussion that follows, examples of an insurance company dental claims processing application will illustrate the impact of the various decisions that must be made when using imaging systems. From this example, applications for other industries, such as banking, brokerage, service, and manufacturing, can be extrapolated.

The criteria for the dental claims application are based on some simple factors:

- Ninety percent of imaging retrieval will occur within thirty days of initial input.
- That same 90 percent must be available within twenty seconds of request.
- All data must be stored for up to three years.

Before delving deeply into the specifics of each criterion, a brief overview of the critical components will provide a good starting point.

### Data Storage Considerations

The imaging system you choose will probably use a combination of storage technologies including disk, tape, and optical media. The choice of one, many, or all of these technologies will determine the capacity and environmental criteria for the system. In general, imaging systems will use a combination of magnetic and optical storage for regular image access. Specific process requirements will determine the need for local or near-line storage.

Using the criteria list for the dental claims application, the high-speed access afforded by magnetic disk storage will provide the responsiveness required for this application. The baseline for disk sizing includes the size of the image to be stored as a factor of the image storage format chosen (TIFF, DDIF, IOCA, etc.), whether or not the images include high-resolution gray scale or color (X-ray and photographic support documentation), and the volume of claims to be processed within a thirty-day period. As will be discussed later, disk sizing will also be affected by the decision for centralized versus distributed system implementation. All this notwithstanding, the sizing of magnetic disk storage is a major decision, as imaging systems quickly consume gigabytes of disk space.

Optical storage and retrieval (OSAR) systems provide economical storage of data for archive and long-term retrieval and analysis. Optical disk systems typically perform more poorly than magnetic storage based on the highly mechanical "jukebox" methods used for disk selection and mounting. Optical systems have the unique advantage of data integrity over extended time frames. Unlike magnetic storage, once an optical disk is encoded with data, the data will not be deleted. Even highly secured magnetic storage degrades over time as the ability of the medium to reliably hold a magnetic charge wanes.

### **Retrieval Times Impact**

Not all issues regarding retrieval are limited to the magnetic and optical subsystems; the overall implementation must be carefully considered. However, a combination of well-meshed technologies is critical to optimal performance. A complete analysis of workstation, network, and storage is necessary to fine-tune the system for maximum responsiveness. Once all components in the system have been integrated, the weak link in the chain is generally the optical subsystem.

### **Optimized Use of Resources**

The most reliable implementations of imaging use this powerful combination of magnetic media for high-speed, short-term storage and optical disk for long-term data management. Many smaller systems can be used on existing LAN systems. Add-on optical subsystems to a variety of servers running an equally wide variety of network operating systems can provide an economical means to provide magnetic, optical, or combo storage systems. Depending upon the demands of your implementation, you may choose to use existing LAN servers for one of many magnetic storage requirements for imaging. These include, but are not limited to, scanning, indexing, and short-term retention.

### **Centralized versus Distributed Criteria**

The nature of a business has a major impact on the choice to centralize or distribute the system. Obviously, a single-location real estate office will centralize. In the



dental claims application, it may or may not be advantageous to distribute certain portions of the system to regional sites. Once again, the complexity of maximizing performance coupled with the need to economize where possible will be greatly impacted by this decision.

### **Legacy Systems Integration**

If your company is like most, you'll not have the luxury to start with a clean slate. Legacy systems from small PC LANs to large global networks will have already been built. Equally large and politically complex fiefdoms to support these systems will provide a variety of supporters and detractors to the use of imaging technology. In most cases, linkage with these legacy systems is necessary to avoid duplication of data already managed and maintained. Although not all components of legacy systems are reusable for imaging, it is generally possible to integrate the absolutely necessary parts through gateways or front-end applications.

---

## **17.2 INTEGRATION WITH CURRENT SYSTEMS**

---

### **PCs, Minis, Mainframes**

Integration with legacy systems is a major challenge. Successful installations use the dynamics of the graphical user interface (GUI) on the PC to build interfaces to mini and mainframe applications. Front ends using GUI application generators buffer the user from the text-based application while concurrently capturing and submitting the entered data and communicating to other applications in the GUI environment. For maximum capability, systems larger than PCs should be LAN attached. Application software for async and 3270 terminal emulation over LANs is readily available. Although not absolutely necessary, LAN access to larger systems provides a more easily installed and maintained physical infrastructure, although nothing precludes individually wired async or sync connections.

In the dental claims application, a claimant requesting information provides a group claim number and Social Security number for identification. The customer service rep enters these data into a GUI dialog box. Iconized on the GUI desktop is a 3270 emulation application that sends the info to a CICS application on the mainframe. The returned data from the mainframe are captured and displayed graphically, with option buttons providing a variety of choices. In this instance, the claimant requests a copy of the original claim form for submission to a secondary insurance carrier. The customer service rep clicks a button, which initiates the address; then other claim information is loaded into a boilerplate cover letter. An image of a claim form is populated with the claim information, and the whole package is sent to an outbound fax server for transmission to the claimant.

### SNA Wide Area Networks

In general, traditional SNA wide area networks (WANs) are not capable of holding up under the rigors of image traffic; nor were SNA networks designed for imaging. This does not mean that the entire SNA network must be abandoned. Many companies with large-scale SNA WANs have also implemented private backbones of 56KB, fractional T1, full T1, and even T3 links. The more traditional portions of this network infrastructure, low-speed multidrop lines, will not be replaceable. The portion that is capable of reuse in a LAN backbone, however, can be reused for imaging.

### Token Ring and Ethernet

Traditionally, image systems require a high bandwidth for reasonable response time and input traffic. Even with the bandwidth available in Ethernet (10MB) and Token Ring (4MB and 16MB), portions of the image system providing high-volume services such as scanning and indexing will be designed into small segments with only a few workstations.

It is advisable that the LAN infrastructure include both Token Ring and Ethernet capability for maximum connectivity. With this philosophy, image components with preference for one medium type or another can be included in your mix. The ability for workstations on one medium to communicate with servers or services on the other will be supported by routers providing the physical translation services and relying upon the upper-layer protocols of the ISO OSI model for communication.

---

## 17.3 NETWORK INFRASTRUCTURE

---

The previous section opened the discussion on the issues of network infrastructure capacity and pointed out that, in many cases, portions of that system may need to be replaced or abandoned. How does one decide what to save and what to discard? And from that which is saved, must it be rearchitected for the new environment? Beginning with the low-speed SNA or async networks, these architectures will be examined, and a migration path toward the more robust networks needed for large-scale imaging will be suggested.

### Low-Speed Network Overhaul

As stated earlier, low-speed networks lack sufficient bandwidth to support the volume of data generated by imaging systems. Low-speed networks, designed for ASCII data or SNA 3270 block mode data streams, are optimized for short bursts of text information, with even shorter acknowledgments. In contrast, an image containing both sides of a personal check at 300 dpi resolution could be as large as



50,000 bytes. Low-speed multidrop SNA networks, or async networks with statistical multiplexors, must be abandoned and rearchitected for imaging traffic. Likewise, the terminal equipment used for the interface with the systems must also be replaced with a workstation capable of providing the rich graphical displays necessary for imaging. The restructuring of the end-user facilities will usually carry the largest budgetary balance. The geographic dispersion of your system and the choice to build a centralized versus distributed system will also factor into the cost of telecommunications in your network.

### Meshed Architecture

Since imaging will be highly reliant upon a LAN/WAN system, a sufficient degree of redundancy and bandwidth availability is necessary in the successful implementation. A meshed architecture will provide maximum capacity. Figure 17.1 represents a traditional SNA network system using modern multiport DSUs on 56KB multidrop lines. How could this network be redesigned for imaging?

Since the mainframe is no longer an integral part of the connectivity, it is removed along with the 3x74 cluster controllers at the two remote sites. In addition, since LANs cannot be multidropped, the two remote offices will receive private links from the network. This means an additional central site DSU must be employed. The redesigned network looks like that shown in Figure 17.2.

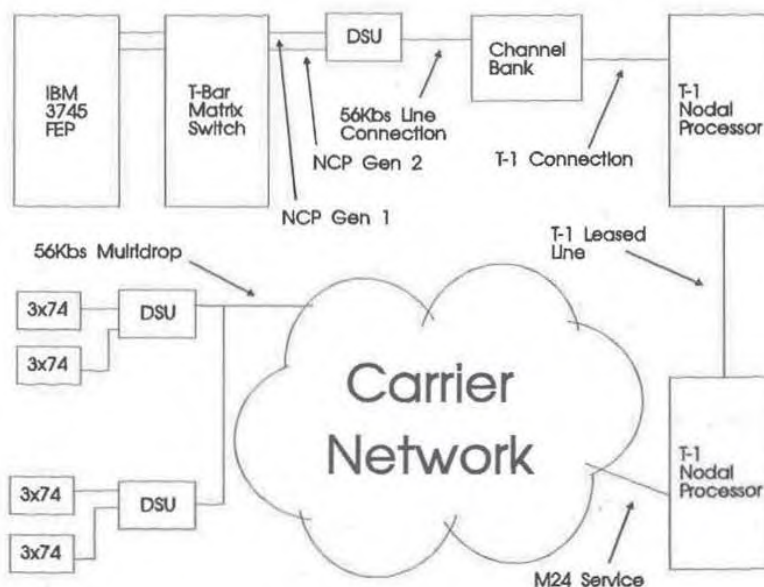


Figure 17.1 Traditional representation of an SNA network

- Much of the SNA WAN has been preserved and used as the new infrastructure for the LAN/WAN needed for imaging.
- Many management systems have also been preserved, thereby reducing training to support the new network.
- Meshing can be accomplished in two ways, as diagrammed in Figure 17.3. One (Figure 17.3a) continues to use T1 nodal processors and channel banks for low-speed LAN/WAN links, and the other (Figure 17.3b) provides direct connectivity into the routers. In either case, the scenario and effect is the same.

The power of this design comes from the ability to recover from the loss of a major leased line. In the meshed nodal processor design, the loss of a single T1 does not disrupt the network, since the nodes allocate and redirect channels over the remaining links. The limitation of this design is its lack of bandwidth load sharing, which is performed dynamically in the meshed router design. All bandwidth in the network is in use continuously until a line loss is detected. Then, redirection of data occurs, such as a meshed T1 nodal system. Many system designs incorporate a combination of both designs in the network. The meshed T1 design allows for the continuation of older systems, which may not have outlived their usefulness.

### Advanced Bridges and Routers

Since the base method for connectivity for imaging systems is local area networks, a decision must be made as to the use of bridges or routers. Using the classic

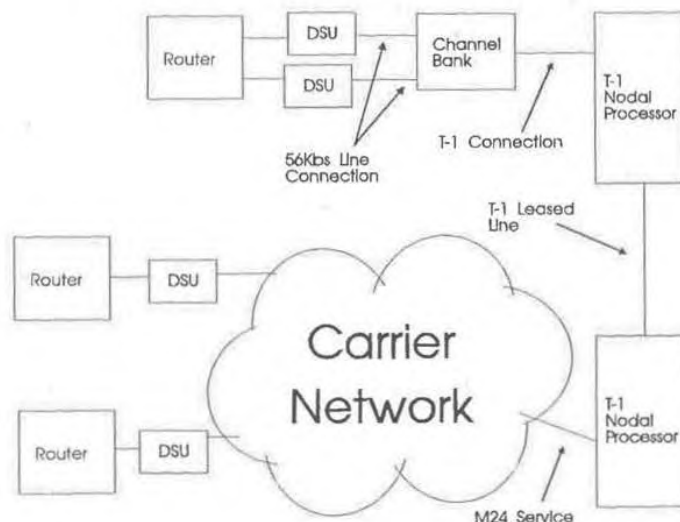
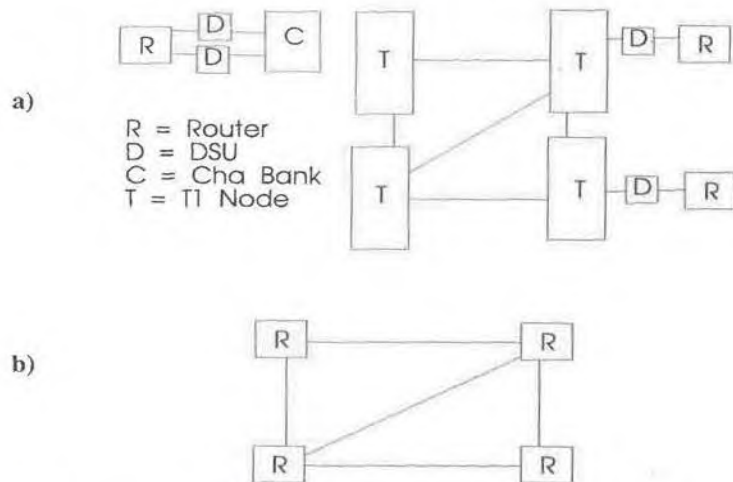


Figure 17.2 Example of a redesigned network





**Figure 17.3** Meshing options: a) meshed T1 network; b) meshed router network

definition, a bridge is a device used to link like networks and a router used for unlike networks. In addition, a router can consolidate multiple simultaneous interfaces, while a bridge is a two-interface device linking LAN A to LAN B. The ability of modern routers to provide all the functions of bridging while also providing interface consolidation and translation between network types makes the choice of routers easy to make.

What brand of router you choose will be based entirely on the complexity of the network implementation. If you are not already familiar with the concepts of TCP/IP and internetwork routing, a considerable learning curve must be accomplished. The power of routing systems allows LAN capabilities to be dispersed in powerful high-speed networks in local geographic areas (campus network) at speeds of 100MB or more and over the WAN at speeds of 1.5MB and up. As WAN technology and the communications infrastructure improves over time, the differences in capacity will diminish.

## Network Management

Like all network resources, the ability to manage the image system components is critical. Tools to analyze capacity and performance, and manage administration and security will typically come directly from the manufacturer. In many cases, software agents that can provide an interface with SNMP-based management systems will provide a reasonable means to provide fault and alarm monitoring. As more advanced management systems based upon ISO or OSF/DME models emerge, a more rich and heterogeneous management environment will be available.

Under these conditions, a performance alarm event may cause the automatic gathering of statistics for analysis and subsequent action by the network designer. Or a capacity trigger may initiate the order for a magnetic or optical subsystem upgrade. Regardless of the mechanism used, an imaging system and its network infrastructure must be manageable to maintain the desired levels of performance and administer to the ongoing capacity needs of the system.

## High-Speed Transports

As imaging capabilities begin to permeate a network, the demands for greater bandwidth will be inevitable. Although high-speed networks for LANs exist today (Ethernet, Token Ring, FDDI), the same is not necessarily true for the WAN. Currently, the most economical communications link is the T1, operating at 1.544MB. T1s are not necessarily inexpensive, but compared to the next level of WAN link operating at 45MB (a T3 line), T1s are the level of connectivity that most companies can afford. Although a company can build a meshed network as described earlier, all the work necessary to manage the data relies on the capabilities of the router. Many companies building LAN/WAN systems are looking to the communications carriers for better WAN capacity. By doing this, the routers can limit themselves to the functions of interface and protocol management, while the WAN handles the high-speed transport of the data.

The carriers are currently developing, and in some cases have limited deployment of, certain high-speed systems that will see greater usage shortly. One of these is frame relay, diagrammed in Figure 17.4.

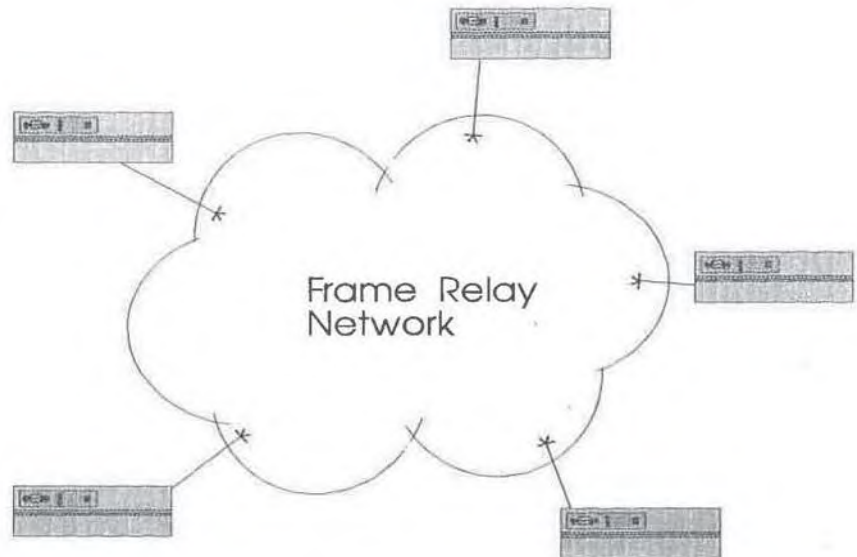
In a frame relay network, the address of each private node attached to the communications cloud is also known to the carrier system. In this way, an addressing function similar to that used on the LAN is used by the carrier to route traffic. In this system, the private and public networks work cooperatively to manage the data and assure accurate, efficient delivery. A frame relay system operates very similar to an X.25 network with a few exceptions. Frame relay operates over very high speed office channel (OC) links running at speeds beginning at T3s 45MB and upwards. This is not to say that X.25 networks cannot operate at these speeds, but the error correction employed by the X.25 network imposes a performance overhead not needed in a modern fiber optics-based system. Frame relay employs the reliability of this fiber network to impose no management overhead. Instead, the integrity of the data must be verified by the private nodes.

Once again, the cooperative nature between the LAN and WAN systems allows high-volume systems such as imaging to effectively incorporate a distributed architecture with minimal impact on system operation.

## Traffic Modeling

Many companies employ tools for traffic modeling to optimize performance. This is very prevalent in SNA networks. Note that older modeling tools may not be





**Figure 17.4** Diagram of a frame relay

upgradable to LAN modeling. However, several good modeling tools can perform LAN, SNA, sync, and async analysis. As with SNA tools, the ability to accumulate accurate traffic data is key. Look for tools that allow the importation of traffic captured by protocol analyzers. Most analyzers provide a utility to export captured data and statistics to an ASCII file. It is necessary to capture data over time and provide the tool with statistically accurate data for good modeling to occur. Without an import feature, your tool is nearly useless, because the data upon which its calculations will be based are raw estimates. An import feature also relieves you from entering pages of statistical data.

---

## 17.4 PERSONNEL ERGONOMICS

---

There's no getting around it: Even the most sophisticated imaging system needs operators to perform key functions. Although the workflow software and process by which the work is done may have changed, the impact on the operators must be taken into account for a successful implementation. This includes not just the ergonomics of the workstation, critical of course, but also the psyche of the operators themselves. In some situations, the radical change brought on by imaging requires adjustments in how people measure their success, and how *they* are measured.

## Workstation Platforms

An obvious necessity for imaging is a workstation with sufficient power and facilities to generate the visual presentation of the data and applications. Questions quickly arise regarding the sizing and scalability of the workstation. Serious consideration must be invested in the workstation decision, as this line item generally represents a large percentage of the image system cost.

For maximum functionality, the chosen platform should be scalable in terms of memory, graphics, and storage. Likewise, the operating system and GUI should allow upward use of the same application software. These decisions will eliminate the need for expensive platform replacement as requirements change and eliminate the need for costly retraining on new software systems.

### *Low Performance*

Low-performance workstations should be given only to the most casual users. These people generally perform intermittent access of images and are under no response-time constraints. Researchers and analysts may fall into this category. These workstations may contain Motorola 68020 or Intel 80286 or 386sx processor types. They may or may not use a software compression application to display the images. Good resolution monitors of VGA level will suffice, although color may not be required. Baseline RAM and hard disk subsystems and minimal network connectivity are provided. These workstations provide the best economy for a wide group of users. Because of the smaller low-resolution monitors, a good zoom facility is necessary to obtain better image readability.

### *Midrange Performance*

Midrange systems apply to those persons with regular access needs who are not in a customer service or highly visible position. Administrators and application developers may fall into this category. Midrange systems typically have Motorola 68030, Intel 80386, Sun SPARC, or Mips R4000 processors and higher. These workstations have high-resolution monitors, often color, at or about the 17" range, better RAM and hard disk facilities, and use software-based compression for faster image display. Although not of the highest resolution, good zooming and access to other peripherals such as modem and fax pools are generally available.

### *High Performance*

These systems are used in areas where maximum performance is needed when imaging is highly visible or customer service is in operation. They generally are based on RISC platforms such as Sun SPARC, Mips R4000, DEC Alpha, HP9000, IBM RS6000, and others. High-resolution 20" monitors with hardware-based



graphics accelerators provide quick display of image graphics. The speed of the processors supports rapid interaction between applications. Due to the size and resolution of the monitors, zooming is not as necessary as with other platforms because the image will be clearly displayed and should be easily read. The personnel impact of using even slightly lower resolution screens when high volume is required should be considered. The result may be frequent breaks for users due to eyestrain and other related symptoms.

### ***Compression Methods***

The method used to display images may be based on the use of software compression and decompression. The decision to use these drivers should be based largely on both the capacity of the workstations and subsystems. It is necessary to work closely with the image system vendor to deploy such systems.

The use of software compression impacts a LAN/WAN system as well as the workstations. At the workstation level, additional CPU and memory resources are needed to generate and present the display. This in turn impacts overall workstation performance and degradation as image retrieval and density increases. Compression has exactly the opposite effect on the network infrastructure. Since the compressed image represents a smaller data record, the network can transport compressed images faster as the amount of data is diminished. In addition, images that are stored as compressed files on the magnetic or optical subsystems consume expensive resources at a lower pace.

In a utopia, images would be compressed at all points in the underlying infrastructure until reaching the workstation, where decompression would occur. As major vendors work to resolve this problem, the ability for "smart images" to traverse the system will become more of a reality. At this stage, however, it is necessary to carefully assess the ability of your chosen vendor's product to work with a variety of software- and hardware-based compression and acceleration at the workstation and storage points.

---

## **17.5 SYSTEM STRUCTURE**

---

One of the greatest decisions you will face in your design is that regarding the centralized or distributed structure of your system. Much of this decision will be based upon your company location(s) and political structure. A small one-location installation will have limited cause for concern; a large insurance company with two or more data centers and regional offices will have a larger task with which to grapple. There are many pros and cons to each approach. Again, a careful examination of your criteria for performance, access, and telecommunications must be made before making this determination.

## Distributed Theory

The choice to distribute an imaging system should be based on a business need for fast response time, optimal network utilization, maximum regional mail service, high disaster recovery, and ability through workflow design to distribute post-scanning activity around the system.

In the dental claims example, a distributed system would provide a means for regional collection of claim information either by mail or fax. Mailed claim forms would be scanned and entered into a queue for indexing and further adjudication. In the pros and cons sections that follow, a discussion of the salient points should help you determine if a distributed system would apply to your plan.

### *Distributed Approach—Cons*

Distributing an imaging system is a more costly implementation. This factor will permeate the distributed system at every level, which is not to say that a distributed system is a poor design, but that the bottom line costs will be greater.

**Higher Systems Cost** Although all components of the distributed system need not be duplicated at each site, many systems components, such as local magnetic storage will be necessary. It may also be the case that a fully distributed system will have multiple scanning, indexing, and optical storage components. This equipment and the staff necessary to perform the various functions will need office housing as well. Even though you may place these users in low-rent suburban office space, the duplication of that office space has a direct real dollar cost.

**Dislocation of Staff** Presuming your infrastructure may currently be centralized, you may experience a dislocation of staff while converting to a distributed system. Even if currently distributed, it may be necessary to dislocate highly trained personnel to various locations during the implementation for local training and installation oversight. Finding personnel to perform this function may require additional staff or using consultants for successful deployment.

**Cost Center Allocation** Many companies are divided by profit and cost centers. The major components of an image system will be expensive and require knowledgeable management. Typically, these support functions will be handled by a centralized information systems group, the cost center of which is likely to differ with the regional locale. Allocation of environmental and other expenses will require negotiation with internal groups.

**Cultural and Political Perturbations** One of the key reasons for distributing a large-scale application such as claims forms processing is the ability to distribute workload among a wide geographic group based on time zones or cost of personnel. The ability to effectively manage important issues among these disparate groups is



more difficult. Credit for work done on nonlocal accounts must be included as part of the overall reward structure to assure continuity of service to customers.

#### ***Distributed Approach—Pros***

The issues that make a distributed system attractive are the high performance and resilience to regional problems or issues. Local databases on magnetic storage will respond rapidly to requests. Since optical storage is typically slow, retrieval across the WAN to a centrally located OSAR will provide reasonable response time. A blizzard in the Northeast does not preclude the claims in their work queue from being distributed to locations in the South.

**Smaller Databases** Since the databases are small and contain only the minimal amount of data needed locally, response times are maximized. LANs run at high speed to make quick work of scanning and indexing functions, database backup, and image retrieval.

**Disaster Recovery** Even though local backup of data will occur, nothing precludes the duplication of the database to a central repository. With this mechanism in place, the loss of a site to disaster allows an alternate site to assume its data and function. A distributed system provides the highest level of system availability.

**Lower Occupancy Rates** Granted, there will be more locations to support, but it's likely that the distribution of the imaging system is in line with a solid business case for distributed sites to sustain your business. For this reason, smaller suburban sites in regional locations will probably reduce site and environmental costs.

**Best Customer Service** Since the database and personnel managing the data are geographically close to the customer, normal business hours for support will apply locally, and data entry problems can be quickly resolved. In general, service to the customer is faster due to the close proximity.

**Mailing Zone Optimization** In those instances where the use of the postal system for original documentation gathering is necessary, short transit times to regional locations will assure that critical data are entered into the system in a timely manner.

#### ***Centralized Theory***

The centralized method for design minimizes overall cost. Collocation of expensive resources, skilled staff, and the necessary environment allows the construction of a system with large capacity. Political and cost center problems are moot, as are issues regarding system management and maintenance. A centralized system represents a single point of failure for the entire network. Alternate network routing, equipment, personnel, off-site storage of backup data, and a strong disaster recovery plan must be considered.

### **Centralized Approach—Cons**

**Disaster Recovery** The negatives of a centralized approach to the construction of an imaging system are not unlike that for any large data system. Should a major system component fail, or the environment housing the imaging system become unusable, the ability to convert to a backup system will include considerable downtime. As an environment becomes increasingly more complex, the cost of disaster recovery services becomes more expensive.

**Delayed Data Entry** Since all input to the system will be performed in a central site, the problems with mail delay will be pervasive. Regardless of where in the world you locate your data center, there will always be a band of mail zones that emanate outward to the fringes of the customer base. Subsequently, customers living on the fringe areas will always be subjected to a delay in service because of the added time to get their input data processed and/or returned. This assumes there is no inbound fax capacity; if input supported by fax is available, then a centralized system has no impact.

**Staffing Needs** A centralized staff may at first seem to be a plus since you will reduce headcount and eliminate redundant training and personnel overhead. But further analysis shows that processing of documents for the business will begin and cease with the daily working hours of the personnel in the data center locale. Thus, a continental U.S. West Coast operation would open for business three hours after the start of business in the East and discontinue operations two hours prior to the end of business in Hawaii. Even with extended hours and additional personnel, there will always be dead periods when service is curtailed.

**Data Error Resolution** Needless to say, humans will occasionally make data entry mistakes. Either the submitter has incorrectly populated a field or the keypunch operator has fat-fingered the keyboard. In any event, the ability to quickly reach the party who made the mistake and correct it will generally be exacerbated by the geographic separation of the parties.

### **Centralized Approach—Pros**

**High Productivity** Since all submissions will converge on a central location, productivity will reflect the high degree of experience of the operators in the business process and methods used to operate the equipment. Expensive high-speed systems can also be purchased to further improve performance and productivity.

**Lowest Hardware Cost** Only the equipment that is required to populate the workers' desktops and provide short-term magnetic storage and optical service need be purchased. The system can be fine-tuned and easily managed due to the tight correlation between system use and capacity. Unlike the distributed approach, each component is used to maximum advantage with little wasted capacity.



**Lowest Telecommunications Costs** There may in fact be no telecommunications cost for a centralized image system unless fax input is promoted by supplying an incoming 800 number. Since the network infrastructure supporting the system is local to the office or campus, telecommunications costs will be minimized. To leverage image capacity, some telecom cost may be incurred for linkage with large customers or suppliers. Even here, agreements that share costs or quid pro quo for storage on a system can limit leased-line expenses.

**Central Storage of Originals** Because all originals are stored centrally, retrieval time, even for paper backup, is reduced to the minimum. A high degree of customer service is assured by centralized storage and processing. It is necessary to include a secure environment for backups and paper documentation. If the documents are legal in nature, there may be no choice but to keep the paper originals; if this is necessary, it is best to use a nearby storage facility.

**Redundancy Eliminated** Any redundancies that are necessary for even the most rudimentary distributed system are eliminated when using the centralized approach. As with any large-scale centralized system, maximum efficiencies and benefits of scale will inure to your benefit. All areas of your implementation will be affected including facilities, systems, and personnel. In many cases, after initial implementation, headcounts are further reduced as the conversion process wanes. The personnel needed to maintain the business process established by the image system is generally about 50 percent of the startup staff. This does not include the staff needed for the infrastructure system.

**Improved Management** Collocation of the image system with other large systems and the support network allows for a tight integration of real-time management and the logistics to provide ongoing maintenance and repair. This will manifest itself in a high degree of system availability and performance.

---

## 17.6 CONCLUSION: THE CHOKES POINTS

---

Throughout this discussion, the pros and cons of each approach, and the highlights and traps of each decision are biased by the general need for performance that meets or exceeds the need to support the business process. It's difficult to be specific regarding the impact of each component decision, because each implementation is unique. The personality of the system you build will be affected by your legacy systems, political perturbations, and a less than optimal budget. But, a few general rules apply to any system.

In real estate, the mantra is "location, location, location." In this endeavor, "planning, planning, planning" will determine how well your implementation performs. Although it's difficult, analyze as best you can the volume of images that may be entering the system, multiplied by the density of the image format you will process. For example, a personal check measuring 2" x 6" comprises 12 square

inches. At a resolution of 200 dots per inch, each square inch needs 40,000 bits ( $200 \times 200 = 40,000$ ) of storage, or 5,000 bytes (40,000/8 bits per byte). Thus, a single side of a personal check will require 60,000 bytes or 58.59KB per side. Assuming that 10,000 checks per day may be processed, a thirty-day magnetic storage system of 18GB will be needed. And that's only to store the images of the front side of checks!!

Likewise, examine closely the throughput of your network system. Carefully analyze the true performance of Ethernet, ARCnet, Token Ring, and/or FDDI systems. The combination of physical system performance and protocol will impact the ability to process the maximum number of images. If these data also flow over a wide area network, make realistic assumptions about the throughput of a WAN link. In general, the greater the bandwidth, the better. Although expensive, be prepared to purchase the fastest links you can afford, and then adjust your application to take best advantage of what you've got. Few companies can afford T3s between sites. But even with a T1 or multiple T1s your decision to use data compression to improve price/performance of the workstations will also help to reduce the demands on the network infrastructure. Capacity-planning tools should be used and programmed to assume the latency delays associated with each component in the network.

If sufficient forethought is exercised, and informed decisions are made about the process to be converted and the methods used to support that process, you will find your implementation of an image system both rewarding and profitable.