

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

- ¹ Belasco, J., *Teaching the Elephant to Dance: Empowering Change in Your Organization* (New York: Crown Publishers), 1990.
- ² Peters, T., and R. Waterman, *In Search of Excellence: Lessons from America's Best-Run Companies* (New York: HarperCollins), 1982.
- ³ 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

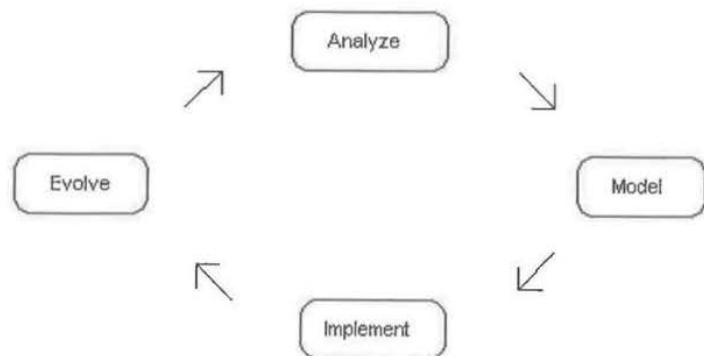


Figure 14.1 The process improvement cycle

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

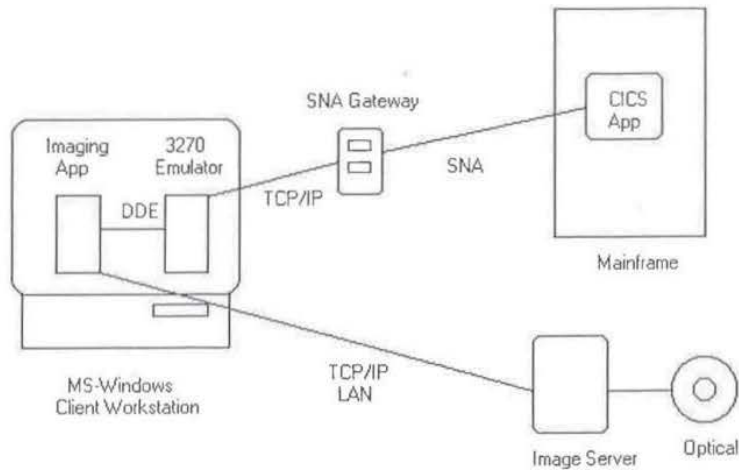


Figure 14.2 Diagram of a 3270 emulation configuration

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.

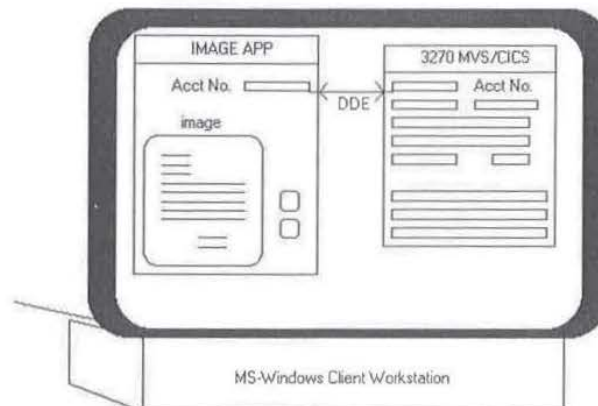


Figure 14.3 Diagram of a DDE screen

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.

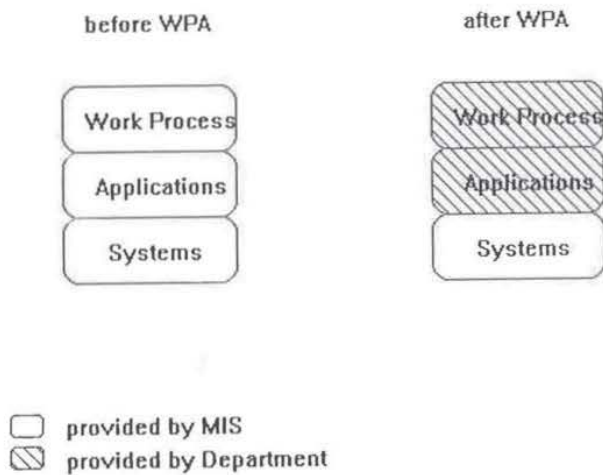


Figure 14.4 Evolution of MIS responsibilities

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.

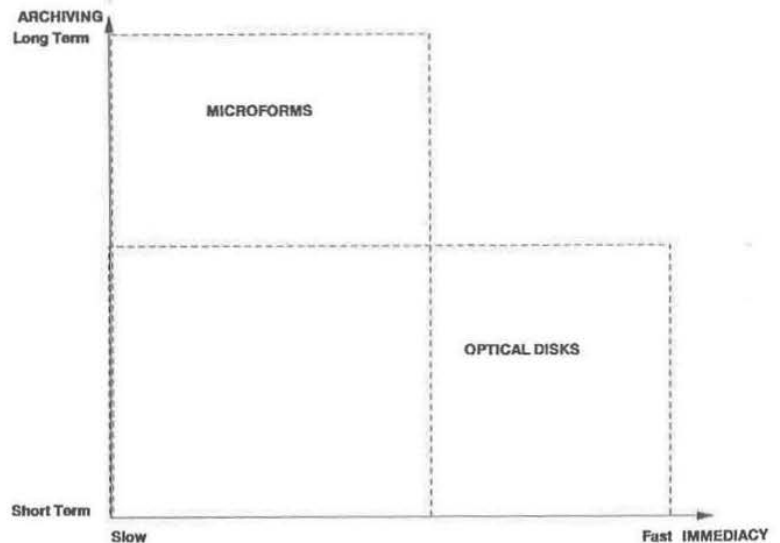


Figure 16.1 Comparison between application areas for optical disk and microforms

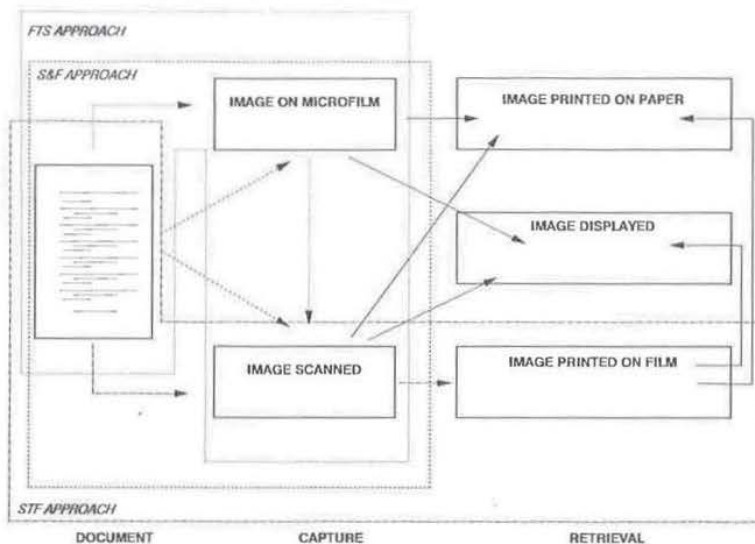


Figure 16.2 Paths for document retrieval

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.

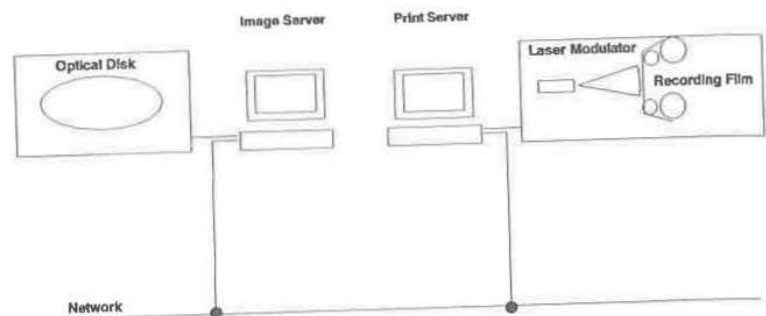


Figure 16.3 Principle of a film recorder

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.

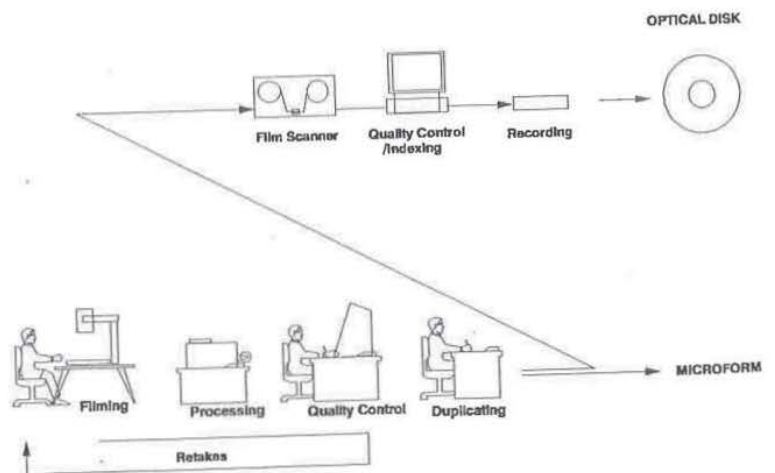


Figure 16.4 Principle of film-then-scan approach

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

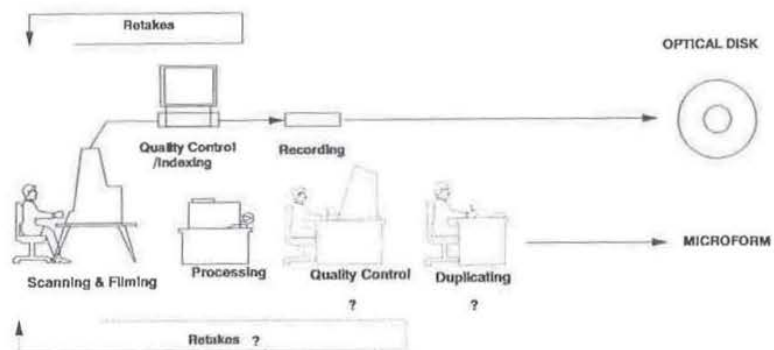


Figure 16.5 Diagram of concurrent scanning and filming approach

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.

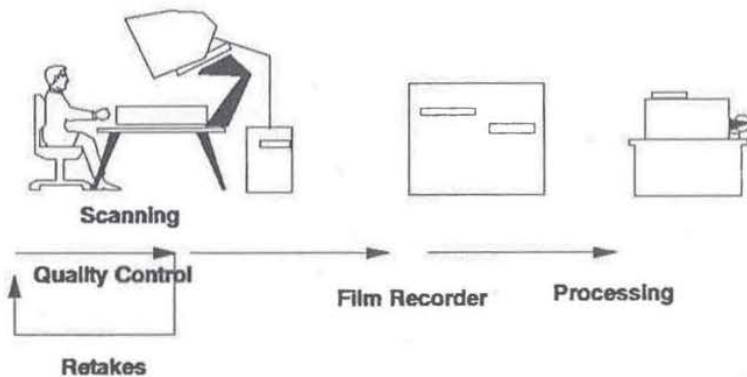


Figure 16.6 Diagram of a scan-then-film operation

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

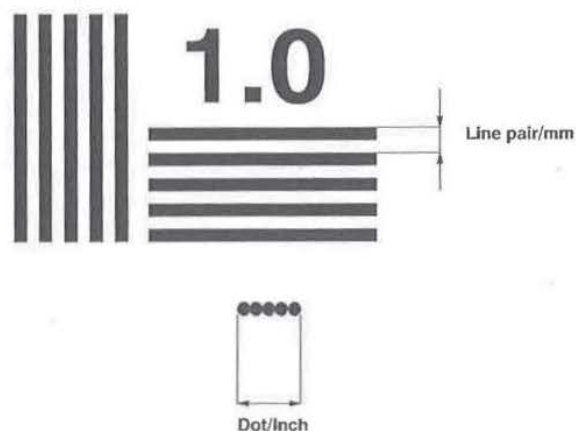


Figure 16.7 Defining image sharpness

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 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

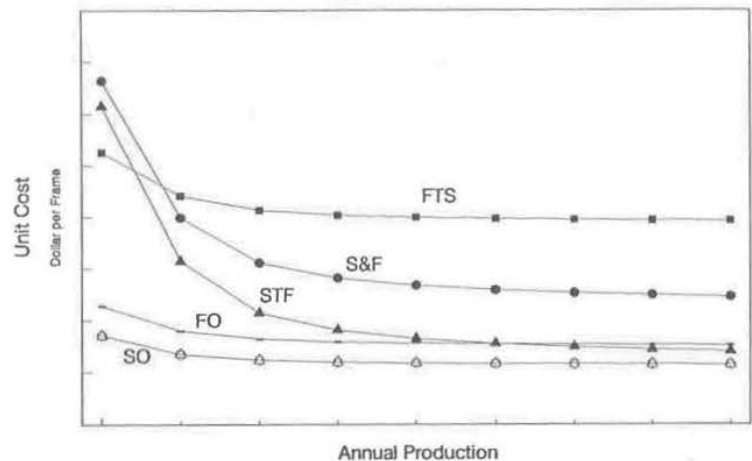


Figure 16.8 Cost comparison between microform and digitized production methods

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&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.

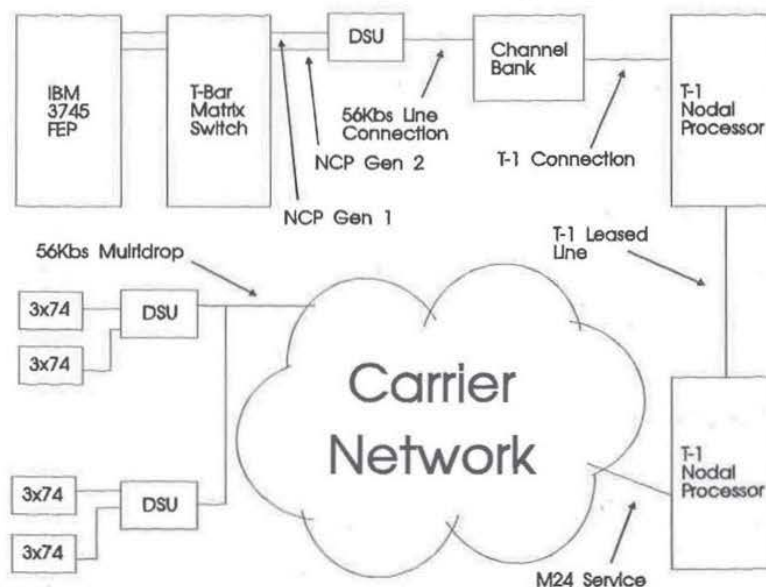


Figure 17.1 Traditional representation of an SNA network

- 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

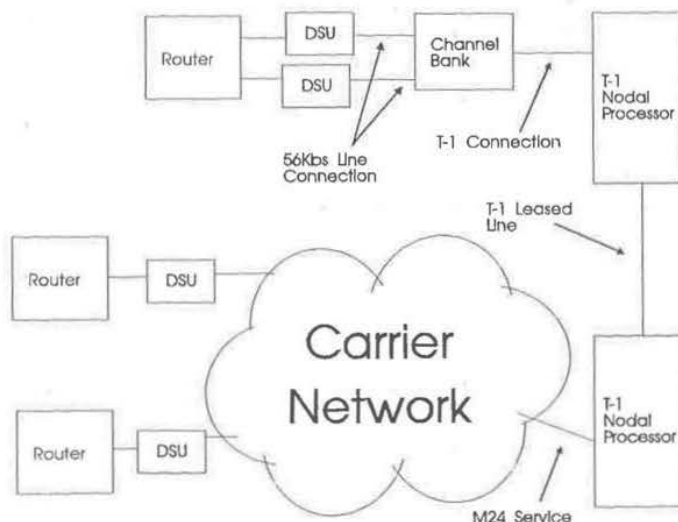


Figure 17.2 Example of a redesigned network

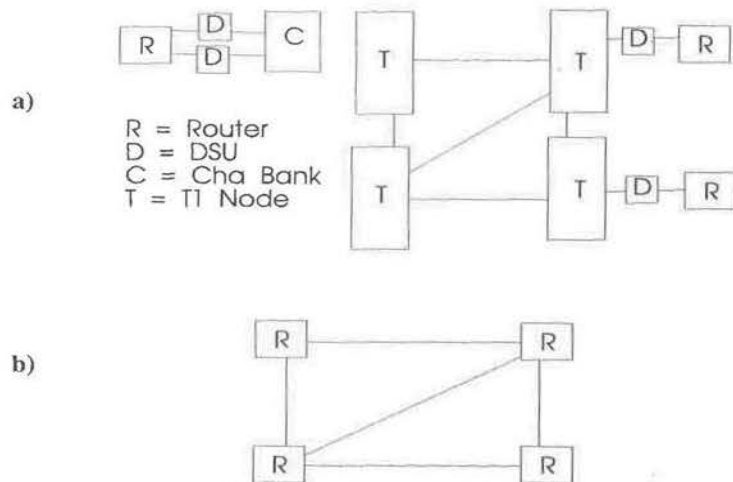


Figure 17.3 Meshing options: a) meshed T1 network; b) meshed router network

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

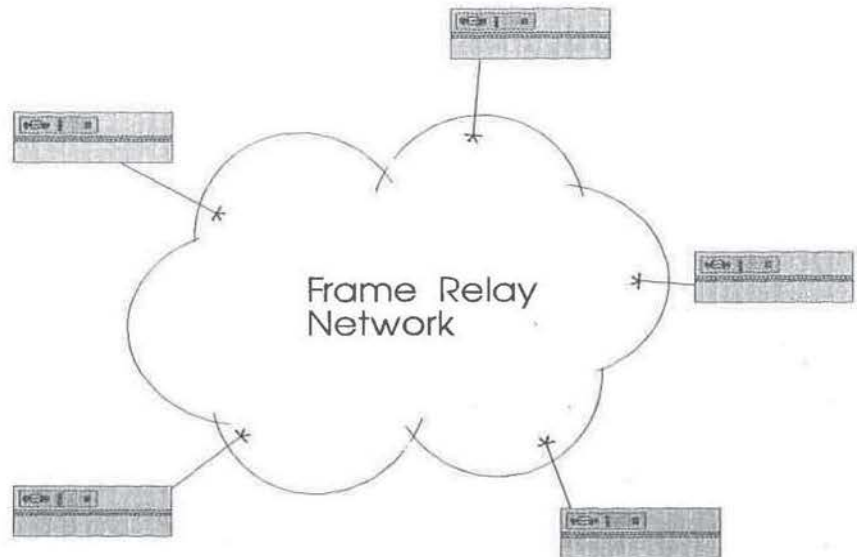


Figure 17.4 Diagram of a frame relay

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.