1998

CPA's guide to document image processing

John Graves

Jacqueline Justice

Marge Cataldo

Kent Information Services

Graves
Justice
Cataldo

The CPA's Guide to Document Image Processing

AICPA

AICPA Top Technology Series

The CPA's Guide to Document Image Processing

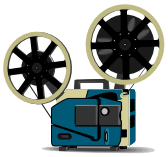
John Graves, CPA
Jacqueline Justice, Ph.D.
Marge Cataldo, M.A.

Published for the AICPA by

● Kent *Information* Services

You can view, print, search, or navigate this book with ***Adobe Acrobat Reader***. To navigate, go to the fully linked Table of Contents and click on an item, or pull down the Tools menu, click on Find, and input a keyword. When connected to the Internet, the URLs in Chapter 8 can be clicked to jump to the referenced site. For more information, click on the movie icon at the right.

AICPA Top Technology Series



The CPA's Guide to Document Image Processing

John Graves, CPA
Jacqueline Justice, Ph.D.
Marge Cataldo, M.A.

Published for the AICPA by

● **Kent**
Information
Services

Title:

THE CPA'S GUIDE TO DOCUMENT IMAGE PROCESSING

Authors:

John Graves, CPA
Jacqueline Justice, Ph.D.
Marge Cataldo, M.A.

Kent Information Services, Inc.
155 N. Water St., Suite 205
Kent, OH 44240
http://www.kentis.com
email@kentis.com
+1 330.673.1300 voice
+1 330.673.6310 fax

Table of Contents

Acknowledgements

We would like to thank the following individuals and organizations who contributed their knowledge, skills, and time to the creation of ***The CPA's Guide to Document Image Processing***:

Scott Stillisano, Webmaster and graphic designer, who is responsible for page layout and design for Kent Information Services, Inc., publications.

Penny Graves and Christa Stalnaker, who coordinate fulfillment and shipping of Kent Information Services, Inc., products.

Lisa Allen, editor, who authors and edits various Kent Information Services, Inc., publications, and who provided the New York Central Mutual case study.

Steve Cembrinski, vice president of operations for New York Central Mutual Fire Insurance Company, who generously shared his expertise and experiences.

Marketing and public relations staff members from Westbrook Technologies, PC DOCS, ISYS, and FileNET, who provided important product demonstrations and allowed us to include useful screen shots of their DIP products.

The American Institute of Certified Public Accountants, who allowed us to reprint the auditing checklist in Chapter Seven.

March 31, 1998
John Graves, CPA
Jacqueline Justice, Ph.D.
Marge Cataldo, M.A.

Chapter One
The Paperless Office

Today's business environment is changing. Nowhere are these technological advances more pronounced than in the area of document management. Experts agree–the move toward Electronic Document Management (EDM) is a natural evolution for any paper-intensive business. Accounting systems, bank loan applications, and insurance claims processors are a few examples of paper-driven enterprises that are exploring EDM in its many forms, including Document Image Processing (DIP).

Even so, the amount of paper generated by businesses today is reaching an all-time high. In fact, there are myriad businesses that specialize in building facilities with one purpose–to store paper documents, pictures, microfilm, microfiche, and other such bits and pieces of valuable information. As we gear up for the millennium, the business that wants to retain its competitive edge must consider alternative methods of managing information and the mountains of paper we produce to create, track, and share information. And accounting professionals must be ready to help their organizations and their clients to face the challenges posed by document management.

By considering the topics presented in each chapter of this text, you will master various concepts related to Document Image Processing technology. Along the way, you will consider the benefits these technologies can offer your organization, and how best to assess your organization's needs, as well as the needs of your clients. But before we explore the details of DIP and DIP systems, let's spend some time considering selected aspects of the

business environment in which this technology is evolving. By way of introduction, we will address the following topics:

- Defining the paper chase.
- BPR and ISO 9000 initiatives.
- Electronic Document Management (EDM).
- Document Image Processing (DIP).
- The role of the CPA.
- Current industry trends.

Defining the Paper Chase

Paper–it's an essential part of our lives. From the smallest offices to the largest, we are often surrounded (if not buried) by reams of paper. We amass stacks of notes, reports, photocopies, printouts, newsletters, faxes, memos, and post-it notes, to name a few. Even in this, the electronic age and the era of the Information Superhighway, our desktops, desk drawers, and filing cabinets are glutted with paper. Why? Well, for one thing, it's human nature to cling to what's familiar. And paper is certainly familiar. It's tangible; we can hold paper in our hands; it requires no special skill or training to use; and it's been around far longer than any technology we can think of. Indeed, the history of paper can be traced to the Egyptians of 5000 years ago. And our individual histories are linked to paper–didn't we all learn to read and write via paper? So, why consider changing our habits? Well, how often have you felt overwhelmed by a mountain of paperwork? How often has vital paperwork been filed incorrectly or simply misplaced? Even when filed correctly, how quickly can you retrieve important documents?

The idea of the paperless office is not a new one; the June 30, 1975, issue of *Business Week* magazine proclaimed that offices implementing (then) state-of-the-art word-processing equipment that would allow users to work on and transfer files electronically were initiating the era of the "paperless office." But rather than reducing the amount of paper generated by a typical office, the advent of the PC has caused a dramatic increase in paper usage. Why? Because nearly every PC used for word processing was (and still is) connected to a printer; we rarely hesitate to create a "hard" copy of whatever is displayed on our computer screens. Obviously, the concept of the paperless office is, for many, not a viable solution. However, for many others, reducing the logjam of paper documents means the difference between maintaining a competitive edge and falling by the wayside.

BPR and ISO 9000

Clearly, the ease with which documents can be created and printed using sophisticated word-processing and printing equipment has played an enormous role in the increase of paper in today's typical business. But, other business initiatives are also contributing to the continuing elusiveness of the long-awaited paperless office. In today's highly competitive market, businesses must keep costs down and profits high. Not incidentally, customer satisfaction must be scrupulously maintained. In short, quality control in any sized business is not a lofty ideal, but a necessity of survival. And the trend towards Business Process Reengineering (BPR) and various other quality initiatives is the result.

Simply put, BPR is the detailed assessment of a business's critical processes. The goal of the assessment is to identify waste and streamline the processes for increased efficiency. Although this would seem to be a process that must, by its nature, identify and eliminate paper waste, in practice, the documentation that is

produced during the redesign process often adds a whole new layer to an organization's existing maze of paperwork.

Similarly, for organizations seeking to gain or to maintain ISO 9000 approval, the creation of documentation for business processes is a necessity. As you know, the ISO 9000 business model is a series of guidelines developed in 1987 and revised in 1994 by the Geneva-based International Organization for Standards for Quality Management. In recent years, ISO 9000 has become the international benchmark among quality systems. So what does ISO 9000 accomplish? Well, the goal of programs like ISO 9000 is to ensure quality control at every step of the production process. And one of the primary tools for maintaining that control is through consistent and detailed documentation of business processes.

You can learn more about the ISO 9000 business model through numerous online sources. For example, the Quality Digest Home Page (**http://www.qualitydigest.com**) provides access to an ISO 9000 database, as well as links to ISO 9000 certified businesses.

Figure 1.1 *The Quality Digest Home Page.*

Electronic Document Management (EDM)

In the midst of this seemingly contradictory mingling of business objectives and business realities, the search for better methods of managing the documents created in businesses continues. In fact, most businesspersons concerned with quality management recognize that one method of reducing costs and increasing profits is to efficiently control the information that is contained in company documents (both paper and electronic documents). According to the AICPA, in the early 1990s nearly "95% of corporate information was still on paper and office workers spent 15-30% of their time searching for information that has been moved, filed or stored." Establishing and maintaining control of these documents must be an integral part of any business's program of quality standards. Electronic Document Management (EDM) is the automated control of page images, spreadsheets, word-processing documents, and complex, compound documents throughout their life cycle within an organization.

Traditionally, there have been two classes of document management: (1) the management of fixed images of pages (printed material, such as books), and (2) the management of editable documents, such as word processing files and spreadsheets. Vendors are rapidly moving away from specializing in one class or another–the best management systems obviously contain elements from both traditional classes. Vendors are now creating larger, integrated document management systems that incorporate a full range of document management functions–these systems control the creation and use of documents across platforms, applications, and company organizational units. The elements of a typical electronic document management system include software to perform all functions necessary to manage the document from inception to archiving. Some of the elements include the following:

- Authoring—These tools support document creation and revision.
- Workflow—Software that allows for the coordination of tasks, data, and people in order to make a business process more efficient and effective.
- Storage—The heart of any EDM system is the database and search engines supporting storage and retrieval of documents.
- Library Services—These are document control mechanisms like checkin, checkout, audit trail, protection/security, and version control.
- Presentation/Distribution Services—Presentation and distribution concerns the form and manner in which users are provided with information. This is a vital feature when businesses wish to reuse information, putting it into a format determined by the target market or business function.

Document Image Processing (DIP)

Reducing the necessary space for document storage has been an issue for decades; and image-processing procedures are certainly not new topics or technologies. Microfilm and microfiche technologies have been available for decades. This technology, in essence, produces a dramatically reduced photograph of the document that is stored on film. They require special machines (called microreaders) to view the stored documents. If a business has an extensive microfilm or microfiche collection, it is typically necessary to employ a librarian to catalog, file, and retrieve microfilm and microfiche documents; and often appointments must be made to view these documents–obviously, this can be a time-consuming process.

Enter DIP technology. A simple definition of DIP could be as follows:

> *A system that digitizes paper documents and stores them as images in some form of data-storage medium.*

Where microfilm and microfiche technologies are, essentially, photographs of documents, DIP creates a digital image of the document that can be stored in any number of ways. And, with the appropriate hardware and software, document retrieval is as simple as the click of the mouse.

The ultimate scope of DIP is the complete management of documents throughout their life cycles from creation, to possible, eventual destruction. The contribution of technology to the capture, storage, indexing, retrieval, processing, and communication of documents is, therefore, crucial to the success of any DIP system. If your company needs to store documents for more than a few months, if those documents are referred to regularly by more than one person, if the storage of those documents requires valuable space and human resources, your organization is a good candidate for a DIP system. (We will discuss the various components of DIP in Chapter Three.)

DIP: A Case Study

To help you visualize how a DIP project may be actualized in your organization or in a client's organization, this section presents a useful case study for a service organization. Although the experiences of this organization do not encapsulate every issue, question, or complication that could affect a DIP project, it is representative of some of what you can expect during a document management redesign project. We hope the stories and

the advice offered by the manager that we interviewed will help you become more confident in undertaking your own project, or in helping clients analyze their own business systems.

New York Central Mutual Fire Insurance Company

New York Central Mutual Fire Insurance Company, currently the fourth largest provider of auto insurance in New York State, holds a century-long tradition as a regional carrier of homeowner, small business, and automobile insurance. Headquartered in Edmeston, New York, the company maintains offices in Amsterdam and Sherbourne, NY, as well as a subsidiary claims adjustment company in Buffalo. Sales are conducted through a network of more than 900 independent, professional agents called Insurance Value Added Network Services (IVANS). With an annual revenue in excess of $350 million, the company has enjoyed dramatic growth over the past decade and now has more than 500,000 policies in force.

Through the early 1990s, New York Central Mutual settled claims the old-fashioned way, using a hands-on, paper-based business system. The company handled about 80,000 new claims per year, and although the average claim file contained 24 pages, more complex claims often grew to several hundred pages. With the manual methods, when a claim came in, the original paper file was pulled from the file room and put on an examiner's desk along with the incoming document. Each day, employees had to sort their mail, prioritize work, sift through the paper file, and decide how to settle claims.

In the 1980s, the company's burgeoning success highlighted the flaws in the paper-based system, which bogged down under the sheer weight of paper necessary to do business. Examiners, who typically handle a caseload of 200 files, spent hours each morning simply deciding where to stack incoming documents

before they could turn to the process of actually settling a claim. Accurately prioritizing, locating, and storing claims information became something of a nightmare. "Too many documents were misfiled or misplaced," said Steve Cembrinski, vice president of operations, who joined the company in 1984. "Under the paper system, employees kept separate files that listed all the missing documents needed to settle a claim, but sometimes, people would lose the 'Can't Find' files. It became a kind of joke. Just the process of keeping files updated was taking up almost 30% of an employee's time."

Answering customer inquiries effectively and efficiently also became a problem. To address a policyholder's question, an examiner would need to physically locate the correct file, then track down information that was either being handled by another caseworker or, worse, had been misplaced. Responding to customers' concerns interfered with daily operations for the examiner and, because of the time it took to answer questions, too often left the customer unsatisfied. Furthermore, file retention and space considerations became an issue. The claims department file room contained about five million documents, with the number rapidly increasing. Under the paper-based system, the company would have had to eventually add storage space.

Another drawback in the paper-based system was the company's growing inability to meet state regulations. The claims process is highly regulated by state agency, with mandated time periods and conditions on the accuracy of settlements. "We were falling behind on medical claims," Cembrinski said, "and had to pay a two-percent penalty each time we failed to meet a target date. This issue, which was a combination of staff and workload, was a problem we definitely needed to solve."

To deal with the rapid growth in the number of policies administered by the company–and therefore in the number of claims to be serviced daily–New York Central Mutual initially added staff, but Cembrinski and the other managers realized the claims process itself needed to be refined. By 1991, Steve Cembrinski began researching work management techniques and document-imaging technology to find a way to better manage business growth. Three years later, in June 1994, a new workflow and document-imaging system, created by FileNET, went online in the claims area of the home office.

Creating Management Goals

In designing their new system, New York Central Mutual's goal was to give the claims adjusters better tools to deal with claims in a timely fashion. Additional benefits would include: more efficient handling of both state regulations and internal deadlines; improved employee morale and customer satisfaction; and a decreased storage burden.

Management, led by Steve Cembrinski, sought to design a claims processing system that would allow examiners to share documents without impeding the flow of work. In any insurance policy, whether automobile, home, or medical, several types of claims are involved–for example, an auto policy may contain claims in the areas of medical, property damage, and liability. While many insurance companies assign one individual to handle every aspect of a claim, New York Central Mutual believes it is in the customer's best interest to create specialty areas. "We have people who are specifically responsible for settling medical claims, for example," said Cembrinski. "Because that's all they do, they become familiar with medical issues, the types of questions to ask, and the kind of results to expect. They can form relationships with doctors and hospitals, for example, and become experts in a specific area. By allowing our people to

specialize, we feel we get the best resolution to the problem for our policyholders."

This practice of having multiple examiners work on a single policy complicated workflow under the old, manual system. The managers responsible for designing a new workflow system were committed to finding a way to allow examiners from different specialty areas to work on a case simultaneously. This led the team to focus on document imaging systems. "With most imaging systems, you have ten or so types of documents," noted Cembrinski. "In our case, we have 800+ different document types, and each holds different priority based on internal and external regulations. We have to prioritize and age each of the documents. Getting people the work they needed to see in a timely fashion was an important goal that could not be reached with any paper-based system."

Once the shortcomings of the old system were identified, the first step toward redesigning the workflow system was to form a project team. According to Steve Cembrinski, having a clearly defined, committed leader was vital to the success of the project. "I was the vision keeper," he said. "If you move into a project like this, you have to have one person who can maintain a hands-on relationship through the entire process and who can pay close attention for a long period of time. When one person is dedicated to the project, it is less likely the ship will stray from course. We wanted a variety of individuals on our team, but we found it was a good idea to maintain constancy of leadership within the project team."

Cembrinski surrounded himself with both "quality members" from the claims community, who had credibility with their managers, and Information Systems specialists. In January 1993, a team of 10 claims department employees and IS professionals

evaluated the company's current work processes with an eye toward redesigning the claims system using document-imaging technology. Cembrinski believed it was essential to have intended users involved in every aspect of the analysis. "It's absolutely key to have the employee's involvement in the process from the beginning," he noted. "After all, it is their eight hour day on that system. They know it best. And, they have to ultimately buy into any new system, or it won't work."

The team looked at which process steps could be changed, eliminated, or combined. "We spent a solid week in which I went through the department with the claims people and learned specifically each of the jobs out there," said Cembrinski. "They taught me the process, and they looked closely at the details of the system they knew. We took various types of occurrences and tried to detail the steps involved in processing, asking question like, 'What happens to that paper? Who else needs to look at it?' We did that for every type of claim."

To better visualize and reorganize steps, the team hung sheets of paper on the walls detailing the settlement process for particular types of claims. "Since we were creating a proposal to convince the senior managers," Cembrinski said, "We kept asking the question, 'With an image system, what officially can we gain?' We found that a significant number of steps in the claims process could be saved–which meant time would be saved."

In a formal proposal, Cembrinski presented the team's findings to senior management, and the project received approval in September 1993. The next step was to research vendors. He attended a seminar hosted by FileNET on imaging systems, but the project team also looked closely at several other vendors. By conducting site visits and viewing demonstrations, the team

narrowed its choices to three vendors. According to Cembrinski, FileNET was chosen on the basis of four criteria:

1. Price.
2. Service.
3. Product.
4. "Gut feeling."

Some of the key factors in the selection of FileNET included: system functionality, a developed customer base, and FileNET's understanding of the high-volume transaction processing required by New York Central Mutual. Although the system was "probably the most expensive," Cembrinski said the project team especially liked the type of long-term relationship it would form with FileNET. Before choosing a vendor, team members had to decide how much responsibility for the new system the company could expect to take on. Should New York Central Mutual enter into a long-term relationship and become dependent on a vendor for any upgrades or future customizing of the claims processing system? Or should it form a short-term relationship and request a transfer of skills that would allow it to take charge of customization and support in-house? To answer these questions, the team had to be honest about both what the claims employees wanted and what the IS professionals could handle. Because the company's roster included a number of highly-trained IS people, the team decided to outsource a lot of the initial application development and provide training so the company could assume upgrades to the system. "We wanted to build at least 10% of this process ourselves," Cembrinski said, "so we could be responsible for maintaining and upgrading it. FileNET helped us do that."

Implementing the System

Only nine months after choosing FileNET, the company saw the first phase of its claims processing system go online. Since replacing paper with desktop PCs was a "cultural change," as Cembrinski commented, the company decided to implement slowly. The home office claims department has 250 employees and is divided into eight territories for auto and two territories for property claims processing. One territory went online at a time. "We implemented them 80%," Cembrinski said, "then went back to the drawing board based on valuable feedback we got from users. Once we got everyone in one territory online, we did a benchmark analysis. Every two months we brought on subsequent territories, so it took almost two years to bring everyone onto the system."

By staggering the implementation scheme, the company, working with FileNET professionals, was able to recognize early on that it would need more storage and higher-level servers than the designers had anticipated. Therefore, the system was brought up to par before subsequent territories were brought online.

By concentrating on one territory of users at a time, the project team also was able to emphasize the importance of training. Many of the examiners felt like "old dogs learning new tricks," Cembrinski said. Some individuals, particularly those 50-60 years of age who'd had little contact with computers, were worried about depending upon PCs as a vital work tool. "We had to figure out the best way to train them on the PC first–and then on the workflow system–so they could feel comfortable and perform their jobs," Cembrinski said. An in-house team of technology experts spent a few days on PC training before dealing with the new system, but soon they realized the employees needed more time and attention. As they rolled out implementation, the training process was refined. "Fortunately, the employees were

receptive," Cembrinski said. "After we got the bugs ironed out, they indicated they didn't want to go back to a paper system."

The FileNET System

The FileNET *WorkFlo* application chosen by New York Central Mutual is used for storage, routing, and prioritization, and integrates with the company's IBM AS400 payment system. With the *WorkFlo* system, incoming documents and claims are scanned, indexed, and written to optical disk. The documents are then categorized into 700 different types, and the *WorkFlo* software automatically matches incoming documents with claim files. Based on document type, the software prioritizes the documents, selects an item for an examiner to work on, and presents it to the assigned examiner's queue. A prefetch of all associated documents from optical disk to magnetic disk is completed for the examiner, which speeds system response time at the workstation. This *WorkFlo* "distributor" is running throughout the day, so that as new items come in they are put in priority order and added to the examiners' work queues. When a telephone inquiry is received, the examiner can make a quick query on the status of any claim in the system.

New York Central Mutual has developed in-house software to do back-end processing tasks using the AS400, such as calculating settlement amounts, setting up for major perils and loss causes, calculating reserves, producing checks, and collecting statistical data for the company's financial needs. *WorkFlo* software uses the Dynamic Data Exchange screen scraping technique to get policy information from the mainframe and presents it to the examiners during the claims process.

In addition, the *WorkFlo* application includes letter writing and automatic faxing with 60 different letter types. A bar code is applied to each outgoing letter. The letter is committed to optical

disk and a pending date is set to wait for a reply. The *WorkFlo* software manages the decision to send a second letter if a reply is not received. If the letter is returned in time by the claimant, it is scanned and auto-indexed by bar code. The *WorkFlo* application puts it into the examiner's queue automatically. The software's ability to manage the pending and rendezvous of incoming information with the original claim file speeds the entire claims process, removing the burden from examiners and supervisors.

Costs and Benefits

In its initial proposal, the project team estimated the cost of implementing the new system at $2.5-2.7 million–but the team aimed a little low. "The area we underestimated was installing the Local Area Network, which we did not have before," said Cembrinski. "We knew we needed one–pretty much everyone doing business right now needs one–but we spent so much time looking at the workflow system and trying to figure out which vendor could offer the kind of system we wanted, that we didn't look closely enough at the cost of the LAN."

After one year of use for 200+ users, Cembrinski estimated the total expense for the system was $3.5 million. But, company management believes the benefits have outweighed those costs. "We think it is paying for itself," Cembrinski said confidently. "We've been measuring it the past couple of years. In terms of things like staffing levels, the number of days it takes to settle claims, and ultimate settlement size, we're very pleased with the results."

According to the company's internal evaluation, the streamlined workflow process has improved the average time it takes to settle claims by two to three days, which means the company no longer pays penalties for missing state-imposed deadlines. Furthermore, customer approval ratings have skyrocketed.

Although no employees were let go, the system has allowed the company to reduce staff numbers by 10-15% through the process of attrition, with more payroll savings expected in the near future. Space savings are also a consideration; New York Central Mutual no longer has the need for a five-million document storage room, so that space has been converted to work space.

The Role of the Accounting Professional

The case study above illustrates the benefits that DIP can bring to organizations and the role of the project manager in implementing DIP. As you know, accounting professionals often encounter the same technology issues as other business managers as they strive to increase workplace efficiency and make the most of dollars spent on equipment and software. And CPAs often are asked by clients to make evaluations and recommendations regarding computer systems and software. As a result, many accounting professionals may one day find themselves in Steve Cembrinski's shoes!

In addition, audits and quality assurance reviews are commonly the responsibility of the CPA, who is frequently the operations, administrative, or accounting manager leading the systems-development project. A basic understanding of DIP is essential to completing an audit of and effectively managing these systems. And EDM systems offer a number of benefits to the financial professional conducting an audit:

- In traditional paper-based environments, evidence and the timing of the audit are constrained by the need to first assemble original documents and all related data (except in situations where microfilm has been used). With EDM,

however, data and documents can be inspected during the regular flow of work, while documents are actually being processed. The data should also be "cleaner" due to the controls over input into the system.

- The auditor can use the system to select and prioritize samples and put them into the auditor's work queue, with all the necessary documents electronically "attached."
- The system can provide a comprehensive audit trail, covering all documents, not only financial data records, indicating which personnel performed which functions with the documents.
- The potential integration of existing and EDM systems (integrating data and paper information) should enable the auditor, with the cooperation of the client, to automate much more of the audit and to "design the audit into the system."
- It is also possible for an auditor to conduct the audit remotely, without disrupting the work of the users/employees.[1]

As transaction-oriented systems built on electronic commerce and electronic data interchange (EDI) models flourish, this role is likely to expand and extend into other departments. Further, professional organizations such as the AICPA recognize that CPAs are entering technology consulting in ever-increasing numbers; indeed, Image Processing and Document Management was number two on the AICPA's 1997 list of Top Ten Technologies that affect CPAs. The *AICPA Online* Web site (**http://www.aicpa.org**) provides valuable information on how technology is affecting today's accounting professional. In fact, the AICPA Online's Information Technology Membership Section is geared solely to enhancing the technical expertise of CPAs.

[1] *AICPA publication, Audit Implications of Electronic Document Management, 1997*

The fact is, you may already be a document management professional by virtue of tasks you have been or are responsible for now. The AIIM (**http://www.aiim.org**) identifies document management professionals as anyone who is responsible for any of the following business functions:

- Manage or distribute repositories of documents dispersed throughout one or more organizations.
- Disseminate information to internal resources as well as clients and suppliers.
- Use technologies such as imaging and forms processing to input and retrieve paper-based documents conveniently.
- Use technologies such as workflow and groupware to manage both transaction-oriented and collaborative functions of documents in an organization.
- Address the non-technical management issues necessary to effectively process your organization's institutional memory.
- Are responsible for integrating imaging, workflow, groupware, optical character recognition, and other technologies, together with realistic standards-compliance, intelligent organization, and management of documents.

How do forward-thinking CPAs prepare to take advantage of DIP opportunities and position themselves as document management professionals? Studying this text is certainly a good place to begin. And, as with other types of specializations, it is important to join and participate in the major associations related to the field. We have already mentioned AIIM, an association that any CPA planning to consult in DIP should explore and consider joining. DMA (Document Management Alliance) is a taskforce responsible for controlling the specifications used in document

management software. DMA members include vendors and developers, as well as end-user organizations. Involvement in this technical control group can go a long way towards developing your credentials and knowledge in the field. You can find out more about DMA membership online at **http://www.aiim.org/dma/**. Xplor, Electronic Document Systems Association, is another important association to consider. Find out more about the association at **http://www.xplor.org/index.html**. Depending on the department in which you work or the types of consulting in which you plan to specialize, you may also find it useful to explore field-specific associations. For example, human resources professionals who deal with document management may find IHRIM (International Association of Human Resources Information Management) a valuable source of training, publications and professional contacts. Learn more about this association at **http://www.ihrim.org**.

Learn More About Current Industry Trends

You've probably heard some of the buzzwords–COLD (Computer Output to Laser Disc), COOL (Computer Output Online), WORM (Write Once, Read Many), to name a few. Not surprisingly, an entire industry has grown to meet the current and future needs of electronic document management in general, and DIP in particular. Indeed, the technology is available for any size business. And often, the Internet can be the best place to get started on your quest for information. For example, Millennial Vision, Inc. (**http://www.mvimvi.com**), vows to "help each client reduce paper flow and increase efficiency through the implementation of COLD, Document Imaging, and Groupware products."

Figure 1.2
Millennial Vision, Inc.'s home page.

Another good online source is MicroAge's *Document Imaging* page (**http://www.docimage.com/nynj_soon.html**). This site provides an excellent overview of EDM, as well as a wealth of background information, and relevant information on products and services.

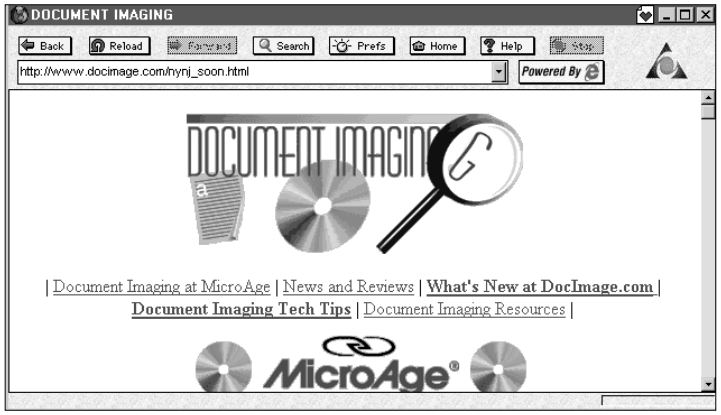
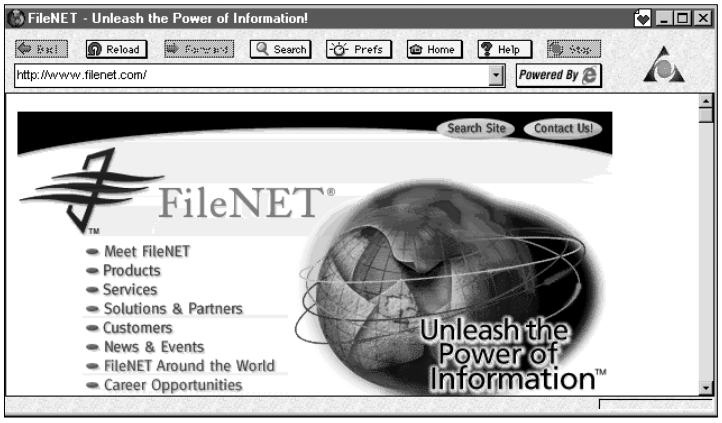


Figure 1.3
The MicroAge Document Imaging Web site.

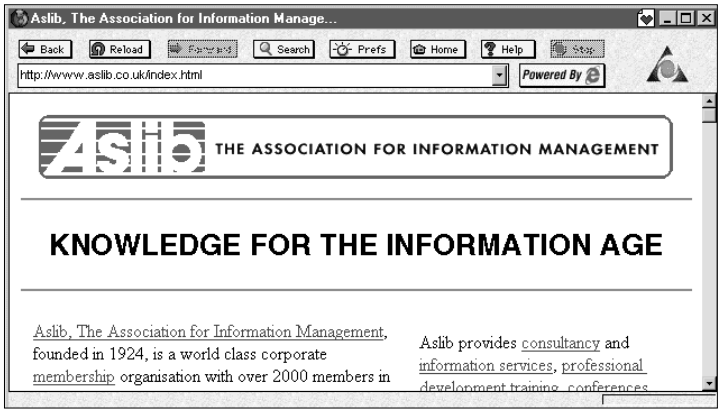
Industry giant FileNET's Web site (**http://www.filenet.com**) is yet another readily accessible site that is a great starting point for the neophyte DIP user. For example, with an excellent internal search engine, FileNET allows you to browse its extensive database, as well as providing access to numerous relevant links.

Figure 1.4 *FileNET's Web site.*



Another useful online resource is Aslib, The Association for Information Management (**http://www.aslib.co.uk/index.html**) where you can access information on membership, training programs, conferences, and numerous links to other relevant sites.

Figure 1.5 *The Association for Information Management Web site.*



The International Information Management Congress (IMC) Web site (**http://www.iimc.org/**) is an excellent resource for keeping current on industry trends. The site provides links to membership information, conferences, as well as an internal search engine that allows you to request relevant information.

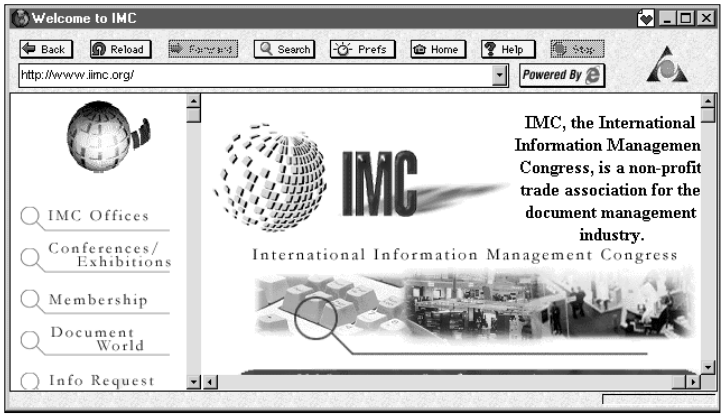


Figure 1.6
The International Information Management Congress Web site.

For a good overview of IT issues and trends, visit the itSMF Web site (**http://exindy.exin.n/ititl/itsmf/**), an independent membership organization. At this site, you can find a wealth of information on industry trends, vendors, and a wide variety of products and services.

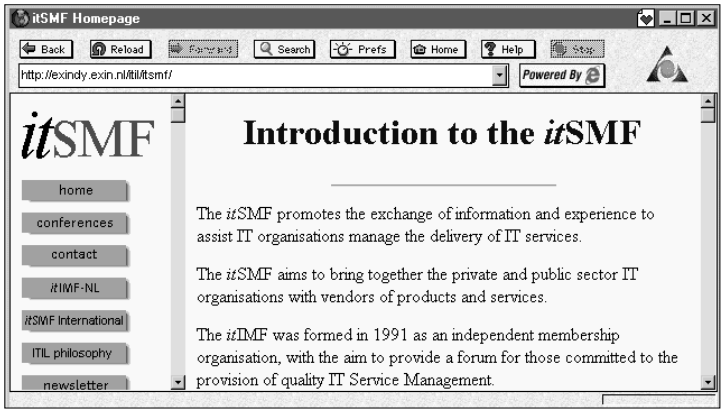


Figure 1.7
The itSMF Web site.

The Web sites listed above will help you get started with your initial research in this exciting information technology area. When you are ready for more, you will find a more extensive listing of resources in Chapter Eight.

Executive Summary

According to Everett C. Johnson, a partner and international director of Deloitte and Touche Enterprise Risk Services and chairman of the AICPA's Electronic Commerce Task Force, "The common thread between information technology and the CPA is *information*. And we've broadened that information beyond financial statement and accounting information to some very sophisticated technology ideas and tools." And, clearly, accounting professionals must master both the technologies that create information and those that manage information efficiently. In fact, in recent years, AICPA members have acknowledged this reality by identifying DIP as one of the top technologies that affect today's CPAs.

Keeping abreast of innovations in any technology can, at times, be quite overwhelming. Today's CPAs and financial managers know all too well that "time is money." How can you sift through the media hype to determine which document management system will prove most beneficial to your organization? Well, the purpose of this text is to provide you, the busy financial professional, with some basic information that will help you make the best possible contributions to the systems development decisions for your organization and for your clients. Whether you are in a small to mid-sized accounting firm or a large corporation, Document Image Processing issues are bound to present themselves. And, since this book provides an overview of the top DIP issues affecting CPAs today, this text can serve as a useful reference tool. With that in mind, let's proceed to Chapter Two, "The Business Case for DIP."

Chapter Two
The Business Case for DIP

In Chapter One, you were introduced to DIP and the business environment in which DIP has become a necessary component of competitive strategy. Even so, the fact is that DIP is not necessary for every organization. And, in many organizations where the benefits of DIP would be most obvious, managers and decision-makers may not recognize the value of this evolving information technology. Accounting professionals involved in DIP planning face two distinct challenges: (1) to decide whether or not DIP (in conjunction with other EDM technologies) is the most cost-efficient way to solve document management problems in their organizations; and (2) to develop detailed business cases to convince managers and other decision-makers to commit the necessary time and resources to implementing a DIP system.

Although we will discuss the particulars of planning a DIP project later in this text, in this chapter, we will focus our attention on the benefits and pitfalls of DIP systems that you must consider when deciding whether or not DIP is the answer for your organization. Topics that we will consider include the following:

- Reduced costs.
- Improved productivity.
- Improved customer service.
- Reduced errors.
- Improved management and control.
- Compatibility with existing systems.
- Various pitfalls and risks.
- Some useful examples.

Typical DIP Benefits

So how can Document Image Processing be applied to your organization? How can you ensure that today's technology investment will not be obsolete tomorrow? After all, it's all well and good for Fortune 500 companies with lush budgets to invest in high-tech gadgetry–but what about the smaller business with more modest resources? And departments with restricted budgets and vast document management needs? At one time, sophisticated document management systems may well have been out of reach for all but the largest organizations, but not anymore. In fact, John F. Mancini, President of AIIM (Association for Information & Image Management International), makes much of this change in a recent report to information and image management professionals (You can read Mancini's report in its entirety by accessing AIIM's online magazine, *Inform*, at **http://www.aiim.org/inform/oct97/Pres1.html**.):

> *With the continuing decline in the costs of our underlying technologies and the extension of the enterprise to the world via the web, more and more users will face document management challenges that are critical to their bottom line. After all, the web is the ultimate document repository. And as more users adopt document management technologies, they will increasingly demand products that are easy to use, implement, and integrate with their existing systems. Our technologies are spreading from our traditional base users with high-volume, departmental and structured document challenges to users with volumes that are more modest, who wish to manage knowledge across the enterprise, and whose "documents" are increasingly ad hoc and unstructured. As document technologies "go mainstream," users will demand products that are as simple to use as the browser that increasingly will reside on every desktop.*

Figure 2.1
AIIM Web site.

Reduced Costs

The most obvious benefit of a DIP system is cost efficiency. Large companies with budgets in the millions, as well as small businesses run from home offices, can both reap the financial benefits of image processing systems. How? Take a few minutes to think of all the ways your organization stores and retrieves information that's crucial to daily business processes. And consider that the documents that would fill several filing cabinets to the brim can be stored on a single CD-ROM. With a DIP system, the amount of physical storage space necessary to support typical business activities is dramatically reduced.

Depending on the size of the organization and the amount of space that is saved, the savings realized from this single result of electronic document storage can take several forms: (1) staff members who previously concentrated on filing and retrieving important documents can be reassigned to more productive tasks; (2) space that was previously used for document storage can be used more efficiently; and (3) the time spent retrieving documents from confusing filing systems can be reduced. Of course, a few moments of thought applied to your own organization will, no doubt, help you to expand this list of potential cost savings.

EVEREN Capital Corporation, a Chicago-based securities brokerage and one of the largest employee-owned companies in the US, found the cost benefits of implementing a DIP system paid off quickly. In the November/December 1997 edition of AIIM's online magazine, *Inform* (**http://www.aiim.org/inform**), Michael Roman, a business analyst for EVEREN, discussed the company's needs.

After establishing the need for a more efficient system to distribute data from various sources to the desktops of myriad employees, EVEREN executives set very clear goals for their organization; they wanted a definite and timely return on their technology investment. Additionally, they outlined the following expectations and needs for the proposed system:

- Enhance customer service.
- Automate processes.
- Reduce the amount of paper generated.
- Eliminate microfilm and microfiche.
- Leverage existing technology by using the current infrastructure and equipment.
- Solid increase in the financial "bottom line."

According to Roman, following in-depth analyses and project planning, the EVEREN team found the system that suited their needs:

> *We selected the COLD solution that was offered by Computron. Our main objective was to use the Computron COOL (Computer Output Online) as our official repository for records retention purposes. We segmented the project into measurable phases, of which the Phase I implementation of COLD was to allow us to distribute mainframe, financial, data warehouse and other reports directly to the end user.*

The results? EVEREN put the COLD solution into production in November, 1996; since then, they have implemented this knowledge-sharing technology in six branch offices with over 275 users. And, according to Roman, "We supply reports from four data sources and on a daily basis integrate over 2,000 reports averaging over 4,000 pages per day. Our growth plans for full implementation is to provide this technology to over 2,000 users in all offices, across all business units." Further, costs are down and productivity is high.

Improved Productivity

Obviously, when all staff members are occupied with tasks that contribute to the business goals of an organization, productivity increases. Unfortunately, in typical business settings, staff members spend a large portion of their workdays tracking down lost, misfiled, or simply hard-to-locate bits of information that they need to complete their daily tasks. In many organizations, these inefficiencies have been accepted as "the nature of the beast." But don't assume that these inefficiencies are irresolvable. Document Image Processing is one strategy for solving these types of problems. A properly designed DIP system places customer files and crucial documents at the fingertips of the staff members who need them, allowing them to complete their jobs more efficiently. And a more efficient staff means additional cost savings for your organization.

OppenheimerFunds Services (the Denver-based division of the financial services company) found a solution for their "paper tiger" through industry giant FileNET. The company's old system operated this way: most of the more than 15,000 daily telephone requests for information had to be submitted to the Research Department where the pertinent items were located on microfilm; then the items were printed, and then the paper copy was forwarded to the responsible telephone service representative

who would then return the customer call. According to FileNET's Customer page on OppenheimerFunds Services (**http://www.filenet.com/cust/oppenheimer.html**), the amount of paperwork was staggering; customers were impatient, and costs were high. The solution was a system called SNAP (Scanned, Networked and Processed) integrated with a workflow application. According to OppenheimerFunds Services Project Manager, Gary Poffenroth, "Since its successful implementation in 1995, [our] document-imaging system has improved our business processes and pioneered a change in the mindset . . . to the point that returning to paper processing is no longer considered a viable option." In short, OppenheimerFunds Services decreased personnel costs by $745,000 and reallocated 10 full-time positions within eight months of the system's implementation.

Improved Customer Service

We've all heard the buzzwords–continuous improvement, quality management, streamlined customer interface; the list goes on and on. However, beneath the jargon lies an important insight. Customers are demanding more, and faster, service for less cost. This is true for all industries, but the professional services are the most clear-cut examples of businesses in which information translates into customer service. We are, in the broadest sense, moving toward a service-based, information economy that will place knowledge at the top of the list of marketable goods. Your organization and the organizations of your clients must be poised to deliver what customers will be expecting–fast, accurate information.

DIP can help many organizations meet these challenges by giving fast information access to staff members who can then meet the needs of clients and customers efficiently and accurately. OppenheimerFunds Services, for example, found that their

customer response time improved dramatically when their DIP system went online–more than 80% of customer questions are now resolved while the customer is on the line; and customer requests are processed in minutes, rather than several days. Obviously, this means happier customers, happier staff members, referrals, and increased profits.

Reduced Errors

One frustration faced by many organizations is the high number of errors that find their way into crucial data. The fact is that whenever manual data entry is part of an organization's process for collecting information, mistakes are bound to occur. Information processing through an automated system reduces the number of errors in an organization's business-critical information sources by using images of original documents in place of databases and files of information manually extracted from those originals. In short, electronic document management removes at least one layer of data input.

In addition, the efficiency of DIP systems has a direct impact on the attitudes of staff members. We all know that nothing succeeds like success. When we can complete a task within a reasonable amount of time, we experience a sense of fulfillment, which leads to a higher confidence level and higher morale. Conversely, when we know from the outset that the assigned task cannot be completed to everyone's satisfaction, we are defeated before we begin. A harried employee, fighting a cumbersome system is more likely to make a fatal error than one who is confident that the task can be completed successfully and in a timely manner.

Improved Management and Control

Obviously, controlling the paper flow also enhances managerial control in any business system. When documents are collected and stored in a systematic way, securing access to selected documents is simply a matter of passwords! On the other hand, making more documents more easily available is also a matter of routine. For example, with a DIP system, gaining control of the "paper tiger" dramatically improved managerial control for Empire Blue Cross/Blue Shield of New York. According to Richard Urban, manager of the Empire Blue Cross/Blue Shield PC Technical Support Group,

> *Paper handling was an issue even for a routine claim, but exceptions further compounded the problem. A customer rep or claims examiner might have to search through a pile of paper to locate a needed document–with a customer waiting all the while. Sometimes, the required document would be at another site, introducing yet another layer of complexity. Paper handling takes time and it's costly–particularly when you consider a volume of 83,000 documents moving through this organization on a daily basis. Paper was the limiting factor in our ability to quickly and accurately process claims and answer customer inquiries. (To read Urban's statement in its entirety, access* ***http://www.kodak.com/aboutkodak/bu/bis/dmc/successStories.html****.)*

In short, availability can be easily restricted or expanded depending on the needs of the organization.

Compatibility with Existing Systems

One final advantage to consider when deciding whether or not DIP is the most feasible solution to your organization's document management challenges is the ease with which DIP integrates

with existing business systems. In essence, the underlying systems that are used to translate facts into useable business information remain the same. DIP simply automates work-intensive, inefficient elements of those processes.

For example, Empire Blue Cross/Blue Shield of New York found that their existing business systems were becoming strained as the volume of work steadily increased. They found their solution with Eastman Software, Inc.'s, *Enterprise Work Management* system. Richard Urban describes the impact of this advantage on his organization in this way:

> *The fact that we can distribute imaging and workflow throughout this geographically dispersed enterprise also has long-term business implications. With the dynamics of today's business environment, we're always looking for ways to run our company better and more cost effectively than the competition. Senior management is continually looking for ways to reduce cost and improve service. [Imaging and workflow technology] give us the flexibility to consider organizational changes and process improvements without uprooting our underlying systems. The imaging system and workflow simply adapt to our needs–however they may evolve in the future.*

Identifying Problems and Solutions

Before continuing, you may wish to spend some time thinking about your organization, the problems that need to be resolved, and how a Document Image Processing system can help you solve these problems. To assist you in this stage of your planning, use the worksheet below to document the problems you identify, your organization's current policies and practices in response to those problems, and the solutions offered by the revised policies and practices that would result from implementing a DIP system.

Identifying Problems and Solutions

Problem	Current Policy or Practice	Solution Offered by DIP

Some Common Pitfalls

In the previous section of this chapter, we focused on the benefits of DIP and shared success stories from real-world organizations that have reaped the benefits of DIP in both improved business processes and bottom-line cost savings. Along with the thousands of success stories, there are always some instances of a system not performing as planned. Of course, as with any business venture, you should exercise caution when you move into unknown territory. But, as the saying goes, forewarned is forearmed. So let's examine some of the most common areas of concern for the businessperson about to embark into the world of DIP.

Planning the Project

No IT project happens overnight and without planning and testing. While this may seem too obvious to discuss, often, the failure of an imaging system is rooted in insufficient planning. We will discuss the details of planning your DIP project later in this text, but before building that plan, be certain that your organization recognizes the importance of this planning stage. Completing the Identifying Problems and Solutions worksheet is a good way to begin. In fact, you may wish to make copies of this sheet to distribute to team members who will be involved in planning your DIP project.

In addition, you and your staff will need to complete detailed costs/benefits analyses and set clear goals with benchmark figures that you can use when testing various solutions to your organization's unique problems. And be ready to ask hard questions when dealing with vendors who are seeking your organization's hard-earned dollars. As you might imagine, competition among DIP vendors is fierce. Therefore, industry

leaders want to convince you that their products are the very best–such is the free enterprise system.

However, this can prove beneficial for the consumer. For example, FileNET offers a free downloadable software program that will help you analyze your DIP needs. Just use your Web browser to access the site at **http://www.filenet.com/prods/index.html** and click on the link for the free cost/benefit analysis tool. Of course, FileNET hopes to generate sales, but that does not lessen the benefit of the analysis tool.

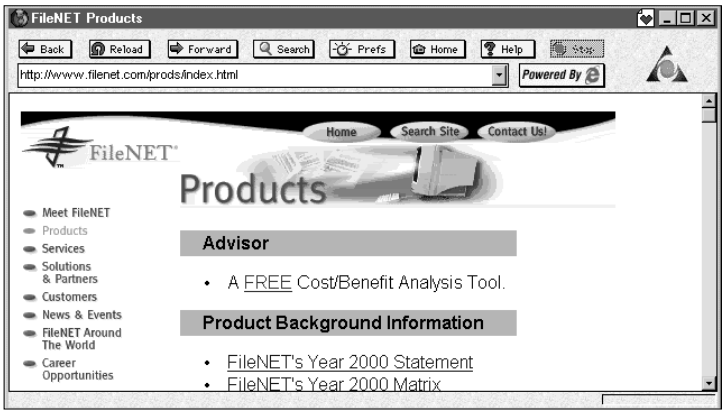


Figure 2.2
FileNET's Products page.

Once the download is complete, the *Advisor* program Wizard will walk you through the setup. Using this program, you can run an assessment of your DIP needs, complete with graphs, charts, and tables.

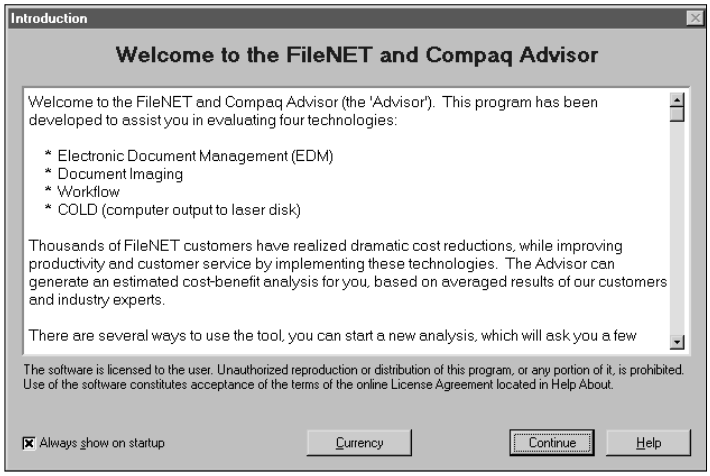


Figure 2.3 *FileNET's Advisor.*

Another free product assessment tool can be accessed through *Document Management Magazine* (**http://infotivity.com/dmindex.htm**). Just click on the link, then enter the relevant information to receive a CD-ROM that will help you take the first steps toward finding a DIP system to meet your organization's needs.

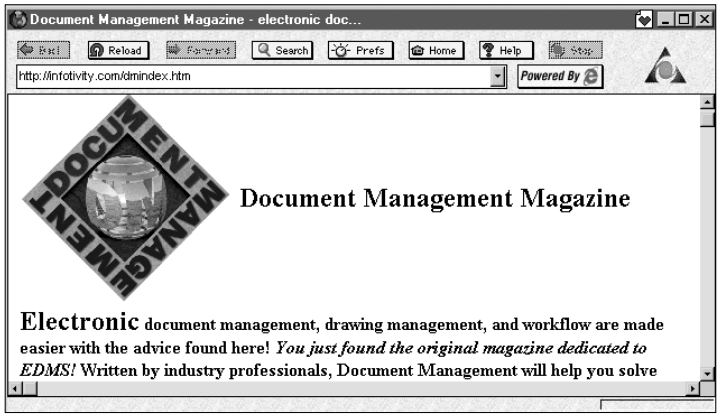


Figure 2.4 *Document Management Magazine's Web site.*

User Friendliness

One of the pitfalls that has frequently led to the failure of DIP (and other IT projects) is the lack of user-friendly software to complete the necessary tasks. Staff members (and managers) who are uncomfortable with new technologies may imagine an office that is suddenly overrun with techno-wizards speaking a language of their own. In this scenario, the rest of the staff look on with varying degrees of resentment and awe–but never daring to venture too near the alien technology. But times have changed. In fact, AIIM president, John F. Mancini (**http://www.aiim.org/inform/oct97/Pres1.html**), cites "ease of use" as a top priority for manufacturers of document management technologies. Further, Mancini cautions members of the industry that while quality was once the main selling point for an information management system, ease of use is now the most important factor in an organization's decision to use a particular technology:

> *Indeed, user friendliness may be to our current age what quality was in the 1980s. In the early 80s, high quality was a differentiator for which consumers were willing to pay a premium. Then the world . . . learned to introduce high quality into their products without significantly increasing price. Suddenly high quality was no longer a differentiator, but a necessary condition of doing business. Those who were slow to learn the lesson paid the price on the revenue line. And that's a lesson that all vendors should keep in mind as our industry moves into the mainstream. Elegant technologies are not enough anymore. We are moving beyond the part of the market cycle where new product features by themselves are enough to succeed. Vendors must focus on using their core technologies and competencies to solve user problems–rather than focusing on technology for technology's sake–or run the risk of being left behind.*

Obviously, the user-friendliness of a DIP system will have a direct affect on the bottom-line analysis of the system's success or failure. If an organization must finance extensive training for a staff to use a system that was intended to reduce costs, the issue of cost efficiency would become moot. Fortunately, today's imaging systems are geared toward a widely diverse group of users; from the company's board members to the mailroom intern, all members of the team must be comfortable with the system for it to be a success. In fact, many systems are being built on Web environments, combining the ease-of-use of intranets with the efficiency of advanced document management and workflow systems. (We will discuss Web-based document-processing systems later in this text.)

Security

Few professions exceed the accountant's in the need for confidentiality. Regardless of the size of your organization, the information generated is at the heart of your success; and that information must be secure from intentional or inadvertent alteration. Planning for appropriate security measures and levels of user rights should be high on your list of priorities if you want your DIP project to be a success. Your imaging system will contain vital company documents that must be preserved and protected against all possible losses. Also, any "down time" of your system will cause customer dissatisfaction–particularly once your customer base becomes accustomed to reaping the benefits of your new system. A good backup plan for hardware and software is always advisable.

Experts who plan for the data security of large mainframe installations use complex terms like "volume dumps," "incremental dumps," "synchronization," and "geometric device differences." These professionals are responsible for planning and implementing a system they hope they never have to use–a

system that will guarantee a major restoration of the contents of many large data storage devices. Each of these data storage devices can hold many gigabytes of information–often the data of organizations such as banks, healthcare facilities, and universities must be stored in such devices. Thousands (even hundreds of thousands) of lives can be affected when disaster strikes such a data center. Obviously, a solid disaster recovery plan is essential. We will discuss recovery plans in greater depth later in this text, but for now, it is sufficient to note that failure to plan for the security and recovery of lost or damaged data can quickly turn a successful DIP project into a nightmare.

Legalities

When considering the implementation of a DIP system, many organizations must consider certain tricky legal issues. For example, can electronic copies of documents be accepted in courts of law and in other legal proceedings? At the heart of this issue is the Federal government's body of law called Rules of Evidence. In short, one item of these laws may be of importance to anyone involved with making DIP decisions. That item states that copies of documents are permissible (as evidence in a court of law) if they can be proved to be authentic in content to the original. The Uniform Photographic Copies of Business and Public Records as Evidence Act puts it this way–copies can be submitted as originals when their authenticity is not in question. This places storage media that can be altered at risk. Storage media (like WORM) that cannot be altered or erased, of course, will carry more weight as valid documents. Considering these issues early in the planning process will help most organizations make choices that will meet their particular needs. And, as with other legal issues, it is wise to consult with a lawyer before problems arise that could interfere with the successful implementation of your DIP system. (Authentication and legal issues will be addressed in more depth later in this text.)

Consult With the Pros

When implementing a DIP system, technology-wise organizations with large IT departments may choose to complete the project with in-house staff members. On the other hand, smaller organizations and technology consultants will frequently find themselves working with outside experts. Consultants and contractors that specialize in designing and implementing DIP systems abound.

It is important to approach these experts well informed and with a lengthy list of questions. Your job as project manager will be to educate yourself about the technology, so that you can communicate the needs of your organization to the professionals who will implement the technology for you. Often, consultants are aligned with particular hardware and software vendors, so be certain to ask about any vendor alignments that will affect the hardware and software choices a contractor will make on your behalf. IT outsourced projects are notorious for the dissatisfaction often expressed by organizations at the project's conclusion. Whether you are the IT consultant hired to implement DIP or the project manager selecting a consultant, much of this disappointment can be avoided by planning, research, and a well-defined purchase contract that clearly states the responsibilities of the organization and the responsibilities of the contractor. (You may wish to consult an attorney who can help you draft a contract that will best protect your organization.)

A wealth of information is available about DIP technologies and the products offered by various vendors. Most vendors even have Web sites where useful information and support materials can be found. (For example, FileNET (**http://www.filenet.com**) and KeyNET (**http://www.keynet.com**) offer extensive technical advice and support at their Web sites, as well as information about products and services.) Although the differences between

DIP vendors and the products they offer can be confusing, current trends toward standardization are hopeful. AIIM (**http://www.aiim.org**) recently announced that the DMA (Document Management Alliance) 1.0 Specification has been unanimously approved by industry leaders such as Xerox, Napersoft, FileNET, Eastman Kodak Software, and Documentum. This can only be good news for the consumer.

Figure 2.5 *Eastman Kodak Business Imaging Systems Web site.*

VARs (Value-Added Resellers)

A value-added reseller can be the imaging novice's best friend. VARs buy the various components for imaging systems and then sell ready-to-use packages (usually customized) for businesses. CompInfo's VARS, Resellers and IT/IS Solution page (**http://www.compinfo.co.uk/cmvars.htm**) is an excellent location to begin searching for a VAR to meet your needs. Like any major investment, taking the time to shop around and compare prices is an essential step of the planning process.

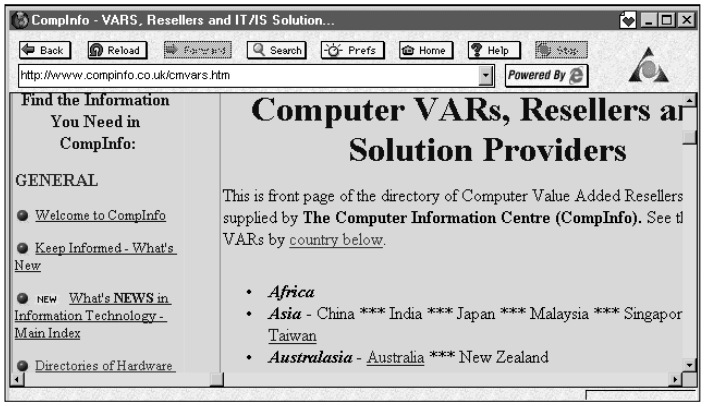


Figure 2.6
The Computer Information Centre Web site.

Avoiding Common Pitfalls

Although it is true that much can be learned through mistakes and failures, the goal of any IT project is to meet the objectives identified in the project plan with as few mistakes and delays as possible. Often, the best way to avoid mistakes is to consider the ways that common problems are likely to surface in your organization. Before continuing, use the worksheet on the following page to consider the common pitfalls described in this chapter in terms of your organization. (Additional space has been provided to allow you to record pitfalls that are not addressed in this chapter, but which may be of particular concern to your organization.) Record the risks faced by your organization and the strategies you could use to avoid them.

Avoiding Common Pitfalls

Common Pitfall	**Relevance to Your Organization**	**Strategies for Avoiding the Pitfall**
Poor planning		
Unfriendly system		
Security		
Legalities		
Outsourcing		
Other		

DIP in Action

As we all know, statistics and manufacturers' hype can be quite daunting, if not downright misleading. But the experiences of those who have "been there" often provide us with the information we need to make the best decisions. With that in mind, let's examine some actual cases.

State of Kansas DMV

Consider the case of the Driver Control Bureau of the Division of Motor Vehicles for the State of Kansas. Although the Bureau uses a mainframe application to maintain records for 1.7 million Kansas drivers, the process of maintaining information on drivers with suspended or revoked licenses was a cumbersome paper-based filing system.

Poor customer service was a common complaint–at one time, the Bureau had a 14-month backlog of cases to process. And, due to the time needed to process telephone requests for information, in January 1992, the Bureau was able to process less than 30% of the attempted telephone calls. Obviously, the system was not an effective one. So, in 1992, the Bureau determined that the time had come to adopt a document management system that would meet the following requirements:

- Reduce processing time.
- Increase internal workflow.
- Increase customer service.
- Provide advanced technology tools for the staff.
- Provide competent management control.
- Enhance employee morale.

The State of Kansas selected FileNET's imaging and workflow products; now incoming documents are scanned, indexed, and routed to Bureau personnel for processing. During processing, electronic images of original documents and driver information stored in the mainframe database can be viewed simultaneously. The results of implementing the system were many and dramatic:

- Reduced the number of manual steps in Driver Control from 27 to 9.
- Eliminated 30 hours per week searching for lost files.
- Reduced average length of telephone inquiry from seven to three minutes.
- Eliminated eleven full time positions saving $227,000 annually.
- Eliminated 143 five-drawer filing cabinets.
- Freed up 1,624 square feet of office and warehouse space.
- Eliminated mundane and repetitive tasks, resulting in higher employee morale.
- Enabled a higher standard of management control.

Learn more about the Kansas Division of Vehicles DIP success story at **http://www.filenet.com**.

Empire Blue Cross/Blue Shield

Empire Blue Cross and Blue Shield operates across 28 counties in eastern New York state and serves nearly five million members. Every day, this member base generates more than 83,000 documents. As a result, paper was the primary source of bottlenecks in processing claims and responding to customer inquiries. Richard Urban, manager of Empire's PC Technical Support Group, puts it this way:

> *Picture tens of thousands of documents circulating among 4,000 people in four separate locations–from the mail room, to a claims examiner or customer rep, and on to storage. Everyone's job depended on paper documents. If someone was waiting for a document to pass from one area to another, the work on the particular item was held up until the paper arrived. Paper handling was an issue even for a routine claim, but exceptions further compounded the problem. A customer rep or claims examiner might have to search through a pile of paper to locate a needed document–with a customer waiting all the while. Sometimes, the required document would be at another site, introducing yet another layer of complexity. Paper handling takes time and it's costly–particularly when you consider a volume of 83,000 documents moving through this organization on a daily basis. Paper was the limiting factor in our ability to quickly and accurately process claims and answer customer inquiries. (Read Urban's account of Empire's problem solving efforts at* ***http://www.kodak.com/aboutkodak/bu/bis/dmc/successStories.html****.)*

Even though Empire transferred finalized paper claims to microfilm, which allowed the original documents to be stored off-site, this did not ease the primary business of processing claims, which still largely depended on paper. And, since a microfilmed document had to be retrieved and printed when needed, bottlenecks continued to hinder efficiency. The company wanted a solution that would improve its fundamental business processes, instead of simply addressing a single element. They opted for Eastman Kodak's *OPEN/image* and *OPEN/ workflow* products; and, according to Urban, the immediate ease of problem areas was remarkable:

With imaging, mailroom employees are the last people to touch paper. All incoming documents go to a scanning station and, from that point forward, the image of the claim or other document is available to anyone who needs it–wherever it is needed. For example, even if the claim comes into the mailroom in Albany, it can be used in our New York office. Where a document is received and where the image is processed has become immaterial. Imaging utilizes a wide area network that links our three major sites, as well as the image index located at our data center in Staten Island. Each site includes its own scanners, archives, servers, and workstations, but a user anywhere on the network can access an image and be part of the process, regardless of where the documents were scanned or stored. Most important, staff at any location can access the information required.

Letters of inquiry are handled in a similar way; they are scanned and routed to a representative or examiner who creates an image folder of all documents relevant to the inquiry for ready access. And the technology can be customized to suit a particular need. As Urban further notes:

We've done a lot to maximize the value of imaging and workflow in our business environment. For instance, we've written custom applications that integrate images with our claims processing system. One such application predicts when and where an image will be needed and pre-stages the image to a particular work area, thus ensuring instantaneous availability to the individual needing it. If a claim is suspended, the claims processing system automatically requests that the image be transferred from the data entry area to the claims correction area. It all happens transparently, in the background. By the time an employee needs it, the image is already there, ready for instant access.

The bottom line for Empire Blue Cross and Blue Shield has dramatically changed. Accuracy and timeliness have improved, resulting in improved customer relations and, ultimately, higher profits for investors. Empire now processes "approximately 90% of the claims within seven days. Prior to imaging and workflow, that was a struggle."

Epson Computer

As their company grew, executives at Epson Computer found an ever-increasing need for a better method of handling customer support. According to John Lang, the company's Executive Vice President, the corporate goal was "to create an electronic library containing technical information on the entire line of Epson Products that could easily and immediately be provided to customers." Working with Westbrook Technologies, Epson chose an imaging and workflow combination that they dubbed "The Epson Connection," which is made up of hundreds of carefully trained Epson employees answering over 80,000 customer calls a month.

According to Lang, members of Epson's customer service staff are now able to "immediately call up visual documentation for any Epson product and walk customers through even the most complex product configurations in an easy, step-by-step approach. And, these same individuals can share this information across the entire organization using the 'Epson Highway,' which integrates *File Magic Plus* with *Lotus Notes* as a transparent layer." (Lang's comments, and other *File Magic* case studies can be accessed at **http://www.filemagic.com**.)

Executive Summary

In recent years, professional organizations, such as the AICPA, have consistently listed DIP as one of the top technologies affecting CPAs, their customers, clients, and organizations. In this chapter, we have explored the most compelling benefits and some common pitfalls that accounting professionals involved in DIP projects can anticipate. Having a solid understanding of these general concepts will prove invaluable to accounting professionals who must participate meaningfully in the planning and implementation of DIP and other electronic document management systems. Once you have identified the risks that might affect your organization and have determined that the advantages of DIP justify a solid investment in time and resources, continue to Chapter Three, where you will learn more about the technologies that make DIP systems work.

Chapter Three
Understanding DIP

As you will remember from Chapters One and Two, DIP is an essential element of any electronic document management system. And the improvement of productivity, quality, responsiveness, and value of administrative operations are some of today's most important business matters. DIP technology supports this mission by capturing documents in a digitized form, and by indexing and storing those images in a manner that ensures efficient retrieval. This chapter will describe the various technologies that are most frequently used in DIP systems to accomplish these goals. Topics covered in this chapter include the following:

- Understanding digitized images.
- Capturing documents.
- Indexing documents.
- Storing documents.
- Retrieving documents.

Understanding Digitized Images

As you know, Document Image Processing refers to the technologies and processes used to transform paper documents into digitized forms, most typically as computer images that can then be processed, viewed, or converted into editable documents. Before examining the most common components of

DIP processes, it may be useful to spend some time learning about the basic qualities and characteristics of digitized images.

As you plan for a DIP project, you will frequently find hardware and software products described in terms of their abilities to create, adjust, and process various aspects of digitized images. Understanding these aspects and their importance to efficiently using images will help you to make informed decisions about the products you may need to purchase later.

Types of Images

Although there are many ways to create computer images and many file formats in which images may be saved, there are really only two basic types of computer images: vector images and raster images. Vector images consist of lines and curves known as vectors. Vector images are most commonly used for graphics that require crisp clear lines, such as graphically created type and logos.

Figure 3.1
Kent Information Services, Inc., logo created as a vector image.

On the other hand, raster images consist of a grid (or raster) of small dots, known as pixels (a conflation of the terms "picture elements"). Raster images are typically used when creating digitized versions of photographs and other scanned documents.

Figure 3.2
Raster image of a coaxial cable created with a scanner.

And, as you might imagine, raster images are the most likely type of image that you will need to manipulate and use in your DIP system.

Resolution

In general, resolution refers to the amount of detail with which an image can be displayed. However, there are various levels of resolution that determine the detail that you see in an image. Some of these resolution measurements are determined in the input stage; that is, the resolution is defined by the way the image is created, the software that is used, and the settings that are selected. For example, image resolution refers to the number of pixels that are displayed in each inch of an image, or ppi (pixels per inch). Typically, image resolution is determined during the input stage and may depend in part on the way the image will be used, file-size restrictions, and the capabilities of the image-creation tools. Obviously, more pixels per inch allows for greater detail, and thus, sharper resolution. But more pixels per inch also results in larger files that may slow your DIP system unnecessarily.

Bit resolution (also referred to as pixel depth) measures the number of bits of stored information that is contained in a pixel. (Remember that a bit is a unit of computer data.) In simplest terms, this measurement determines how much color information is available for each pixel in the image files. More bits (or greater pixel depth) results in more available colors, and thus, more detail (and also, larger files). The choices you make when determining acceptable bit resolutions for images, like image resolution, will depend on how the image will be used and on the capabilities of your system.

Other types of resolution measurements are used when discussing the quality of output–on monitor displays and printouts. For example, monitor resolution is a measurement that is used to define the number of pixels (ppi), or dots (dpi), that can be displayed on a computer screen. Obviously, if the monitor resolution is lower than the resolution of the image, the monitor will only be able to display a part of the image onscreen. In effect, monitor resolution determines the size of the image's onscreen display. Output resolution, on the other hand, refers to the number of dots per inch that an output device, such as a printer, is able to produce. Typically, high-end laser printers produce output of high resolution (between 300-600 dpi), while ink-jet and bubble-jet printers produce lower resolution printouts.

Obviously, it is important to make some decisions about the ways you will be using images on your system, the quality of resolution your staff members will need to work efficiently with those images, and the amount of storage space and memory that will be available. Using today's sophisticated hardware and software, it is easy to create high-quality images with remarkable detail and resolution. But these images require large amounts of storage space and memory, and they require more time to display and print. When developing your DIP system, determining guidelines

for image resolution will be a process of compromising between (1) creating image files that provide the necessary data and (2) keeping the demands that these images place on your system to a minimum.

Color

Depending on the types of documents that you will be processing on your DIP system, you may also need to understand the impact of color images. In some DIP systems, color images are never used at all. For example, if you only need to scan in particular forms or applications that are then retrieved for customer service responses, black and white images may be all your system requires. On the other hand, many DIP systems use color images to various degrees. For example, organizations that use color-coded forms may choose to include color coding in the electronic versions of those forms. Other organizations using DIP systems, may even have sophisticated color graphics (such as product models or research prototypes) among the files that are included on their DIP systems. Whatever your DIP system color needs might be, the discussion provided below will help you understand color in digitized images.

As you may remember from your high school science classes, the human eye recognizes color based on the wavelength of light that reaches the eye. Typically, colors are described in terms of three primary characteristics: hue, saturation, and brightness. In simplest terms, hue is the wavelength of light reflected from an object and is identified by the name of the color. Saturation is the strength (or purity) of the color in terms of the amount of gray that is also blended in the hue. Brightness refers to the relative brightness or darkness of the color.

When using color images, you are, in essence, "fooling" the human eye into seeing artificially created colors. Colors that appear on computer screens, for example, are created by mixing red, green, and blue light from tiny phosphor (light-emitting) dots. On the other hand, colors on paper are usually specified as mixtures of four ink colors: cyan, magenta, yellow, and black (known as the CMYK model). The colors for these printed representations are mixed differently than the colors created with phosphor dots; as a result, color printouts frequently differ from their onscreen counterparts. To reduce these differences, you may need to calibrate your DIP system. Calibration is the process of adjusting image values throughout your DIP system (scanners, monitors, software, and printers) to ensure that each link in the DIP chain of events represents colors in the same way. Although, it is unlikely you will ever achieve a perfect match, careful calibration can go a long way towards the control of color images in your DIP system.

If you are considering the use of color images in your DIP system, you must carefully balance the benefits and drawbacks. Like high-resolution images, color images place a greater strain on your system and may require investments in higher-priced hardware and software to effectively manage and control.

Overview of DIP Systems

Every organization will bring unique needs and challenges to a DIP project. Even so, to successfully plan and implement a DIP system, it is important to have a general understanding of the technologies and tools that support DIP in typical organizations. Some organizations implement only some of these steps in-house, outsourcing others to service companies. Other organizations implement DIP systems that complete the entire DIP process in-house. The needs and budget of your organization will determine how much of the DIP system technologies described below will prove cost efficient for your DIP project. In the sections that follow, we will overview DIP systems and provide basic descriptions of the four most common components of a DIP system:

1. Capturing documents in an electronic form.
2. Indexing the documents for easy retrieval.
3. Storing the documents.
4. Retrieving and outputting the documents.

The diagram on the following page illustrates the relationship between these components in a system that uses all four steps:

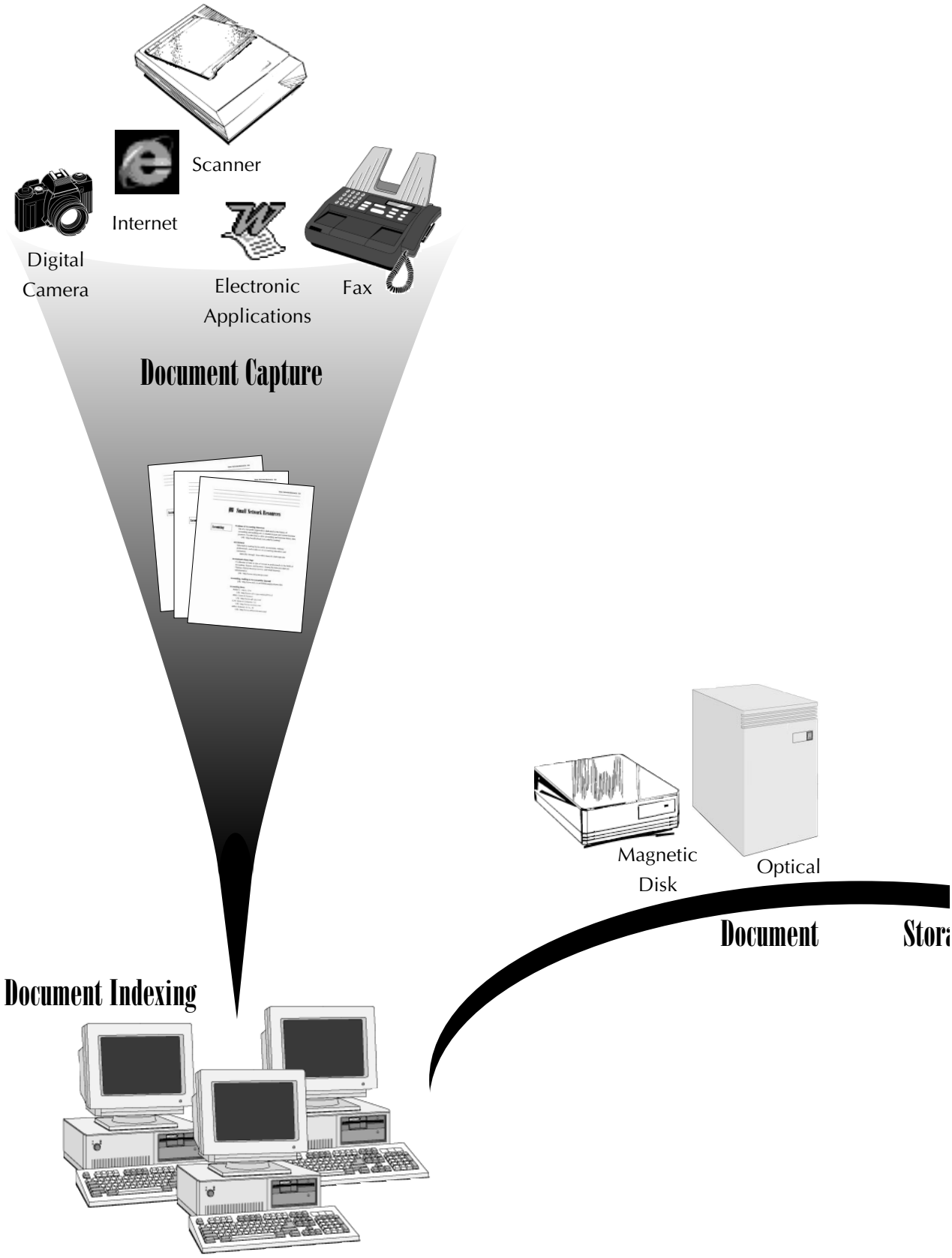
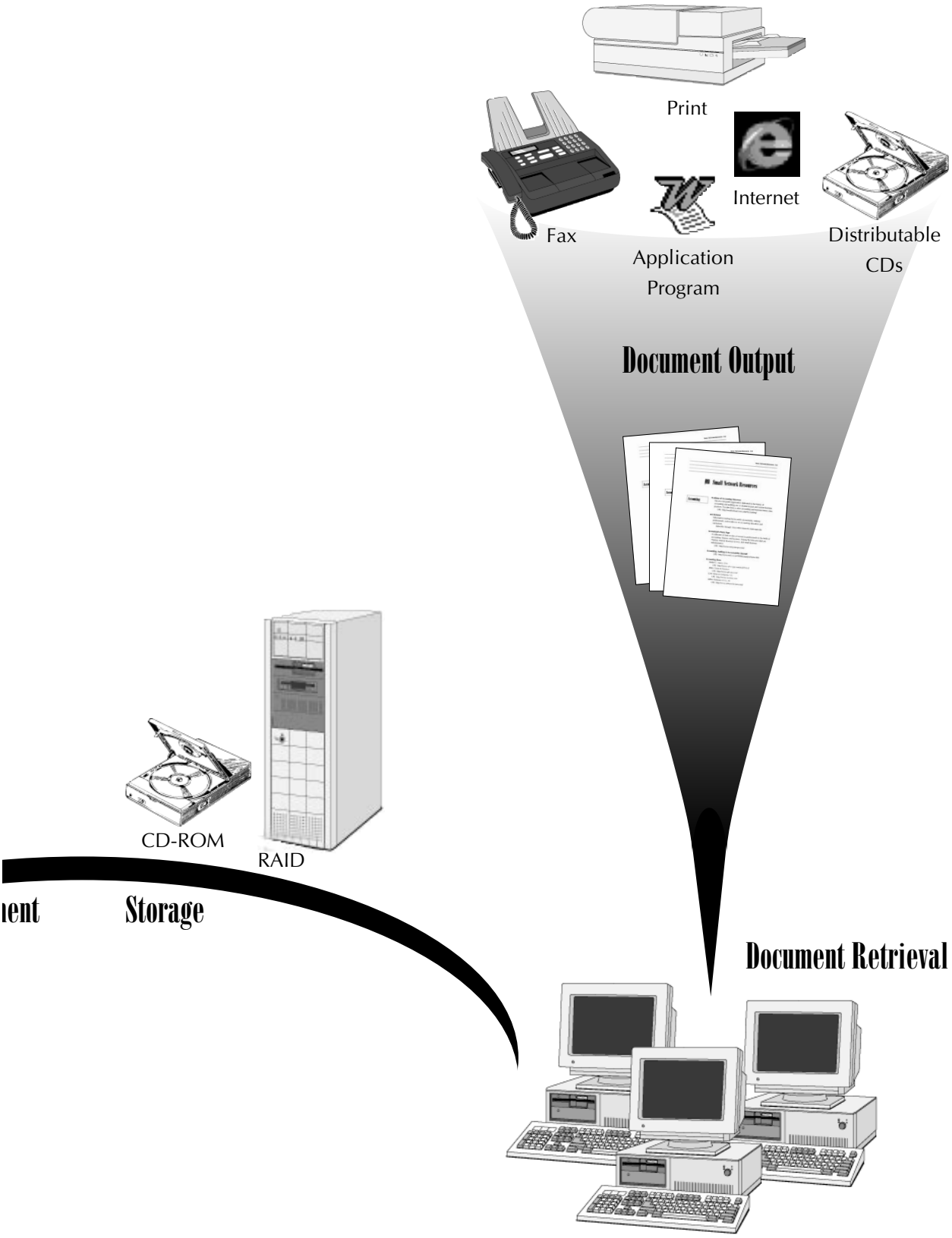


Figure 3.3

The four primary components of a DIP system.


Print
Fax
Internet
Application Program
Distributable CDs
Document Output
CD-ROM
RAID
ıent
Storage
Document Retrieval

Capturing Documents

When developing a system in which most or all business-critical documents are available in electronic forms, you may need to consider a number of separate business processes. Many documents created in organizations today are already created in electronic forms–such as word-processed documents, e-mail messages, spreadsheets, etc. Unfortunately, in many organizations, the final step in the creation of these types of documents is the printing of a hard copy that is then filed and used by other staff members. Obviously, your EDM system should include clear guidelines for making these documents electronically available to staff members who will need them. In fact, you may find that a large portion of your archive (where your digitized documents are stored) will be built from these types of common electronic documents.

Depending on the strategies that your company has used to store documents in the past, you may also need to plan for the inclusion of microfilm and microfiche in your system. If you are planning to bring an existing microfilm/microfiche library online, you must consider the special scanning requirements for these types of documents when selecting scanning hardware and software.

Other types of processes for which you must plan are those that happen offline–forms that are completed by customers and clients, correspondence that arrives through postal mail, checks that are sent by customers, etc. These processes typically result in paper products that are needed by various staff members at various times. The most common way to make these types of documents a part of your organization's electronic archive is through scanning, a process that we will describe in the section that follows.

Scanning

In a DIP system, images of various types of information (paper documents, graphics, transparencies, film, etc.) can be captured by digital scanning. The machines that scan these objects are known as scanners. Scanners use a light-sensitive device, called a charged-coupled device (CCD), to convert light that is reflected from an object into an electronic signal. Simply put, the scanner "reads" the light and dark areas of the page and determines whether a given point on the page should be represented by a black dot or a white dot. (More precise scanners recognize various shades of gray and color scanners use red, blue, and green filters to identify colors in the light reflected from the scanned object.) These dots, or pixels, are stored electronically as an array of brightness values, or a bit-mapped image of the original page. This image can be retrieved and printed or displayed, but in this form it is not editable text–it is simply a raster image.

Scanners come in various types, such as flatbed scanners, which look very much like photo copying machines. When using a flatbed scanner, an object is placed on a transparent surface, the CCD moves under the original document to collect the data it needs to create an electronic representation of the object. Other types of scanners, such as sheet-fed, use rollers to move the original object past the fixed CCD. Handheld scanners allow users to manually move the CCD over an object. While convenient for traveling staff members, hand-held scanners cannot easily scan in the full-page objects that are most typical in business environments.

In addition to the hardware described above, scanners typically require special software. One component of the software is used to control the scanning process, another is used to adjust the appearance of the image, and a third may be used to recognize

text in the scanned object. This software allows you to make important choices that will affect the quality of the image created and ways that image will be used. For example, you will be able to control the image resolution, which, as you know, determines the detail with which the image will be displayed. Your scan resolution choice will depend on how the image will be used. For example, if the scanned object will only be displayed on a monitor, the resolution need only be as good as the monitor resolution can accommodate. On the other hand, if the scanned object will be printed, the resolution that your printer is capable of supporting may determine the scan resolution. Or, if you will be using an automated process to convert scanned documents back into editable text, you may need a high-resolution image in order to reduce the number of conversion errors. The screen below shows an example of the settings that may appear in scanning software.

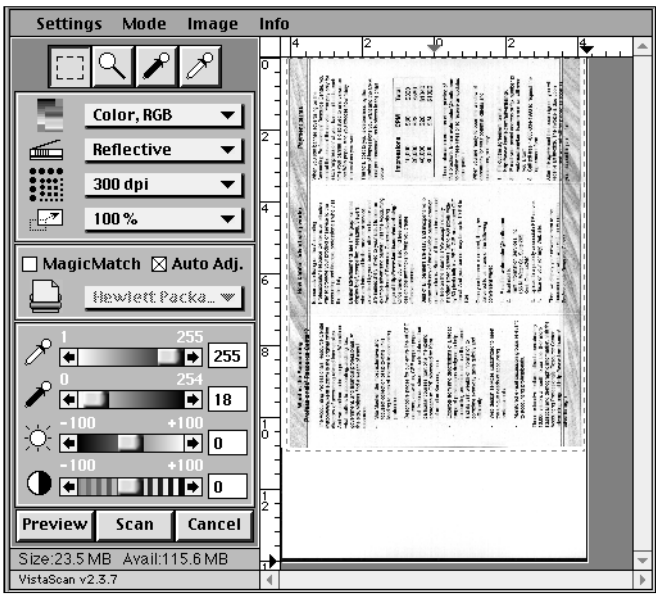


Figure 3.3 *Determining scanner settings in VistaScan scanning software.*

This screen is divided into two panes. On the right, the scanned image is displayed (in this case, a brochure). On the left, the various settings that can be used to control the quality of the scanned image are displayed. For example, these settings allow users to select the types of images they wish to create (such as Color, RGB), the type of document from which the image is being created (such as Reflective, used for paper documents), the resolution (measured in dots per inch), and the scale (measured in percentages of the original document). Other image qualities can also be adjusted, such as highlighting, shadows, brightness, and contrasts. (Vendors of scanners and scanning software will be discussed in more depth in Chapter Four.)

OCR/ICR

In many DIP systems, images that are scanned must also be converted into data that can once again be processed by computers. One of the processes typically used to implement this conversion is OCR (Optical Character Recognition). Most typically, this conversion step is handled as a step in the scanning process. In fact, many scanning software suites include some type of character-recognition conversion utility. A variety of packages are available which run under *Windows*, some of which even recognize handwriting and signatures! Simply put, OCR software enables a computer to "read" the words on an image and convert elements of that image back into text that can be processed by a computer.

Essentially, the images of fonts that appear in the document are compared to a series of standard font images. The OCR software simply looks for a match and then converts the recognized character into ASCII (simple text) form. The text can then be exported into other applications, such as database programs or word processors, where the data can be processed and used in whatever ways the business processes of an organization require.

Obviously, OCR technology dramatically reduces the amount of data entry needed for data-processing applications by creating a bridge between images and other applications in a DIP system. (OCR may also play an important role in the indexing of documents. We will describe this in more depth later in this chapter.)

To better understand OCR, study the five steps that might be performed in a typical OCR conversion process, listed below:

1. The software examines the bitmap image created by the scanner and uses the white space on the page to identify headers, paragraphs, columns, and graphics.
2. The software looks for exact matches between the pixels of scanned images of characters and the pixels of the character sets stored in the program's memory.
3. For characters that are not immediately recognized by the program, the software might analyze features of the characters (such as straight lines, curves, and loops) and then build an alphabet of characters based on this feature analysis. Subsequently, the software can use this alphabet to identify characters that were not converted in the initial character match.
4. Any characters that are still unconverted are represented as a symbol (such as "@"), so that a word processor can be used to manually search for unknown characters and replace them with appropriate character choices.
5. Finally, the converted image can be saved as an ASCII file or in some popular word-processor format.

In addition to the strategies described above, some OCR software packages use spell-checking features to locate obvious errors and to provide a list of probable alternatives for words with unknown characters.

ICR (Intelligent Character Recognition) extends the usefulness of OCR by using sets of rules that correspond to font characters rather than particular sets of fonts. For example, in ICR conversion systems, certain rules can be used to define the general appearance of the letter "T." These rules will be equally applicable to any style of font. ICR software can even recognize different font styles used within a single document.

Both OCR and ICR can be implemented through independent hardware or software that works with the image scanners used in your DIP system. The hardware implementation may be considerably more expensive, but volume processing is significantly faster. Typically, high-end systems are expensive, but they provide higher throughput rates and allow recognition steps to be completed in the background without tying up other system workstations. Because of their low cost and flexibility, the software systems have become popular in situations where high processing speeds are not required.

PDF (Portable Document Format)

One additional strategy for capturing data that we will overview in this section is the use of PDL (Page Definition Language) to capture layout and format data for documents created in various applications. One of the most popular PDLs is PDF (Portable Document Format); in fact, the IRS makes tax forms and instructions available for free download from their Internet site as PDF files. When an electronic document is created, the application program embeds various codes and data in that document, which define the way the document will look when it is printed. When the document is sent to the printer, it is this information that is used to create a printed copy.

PDF simply captures that data and uses it to create an electronic document that looks the way it would look in the application in which it was created. However, the document can be viewed, with all of its unique formatting in tact, in a program called *Adobe Acrobat Reader*. This means, for example, that product lists created in *WordPerfect* by a vendor with whom you do business can be viewed in their fully formatted form even if you do not have *WordPerfect* on your machine. And PDF files are searchable!

The potential benefits of PDF in your EDM plans is obvious–PDF allows you to make different document formats available on your system so that all users can access all information, without the delay of a scanning or conversion process. PDF is best described by the cofounder and CEO of Adobe Systems, Inc., John Warnock:

> *The history of PDF started in 1991, when it occurred to me that everybody has been talking about a "Paperless Office" for decades, and we envisioned it all as plain ASCII green letters, all uppercase. The "Paperless Office" has always been the Holy Grail, and I looked around and saw computers on top of all the desks. This meant we could grab all the content as printer output, and this was critical to solve the font problem.*
>
> *The three seed inventions that led to Acrobat were the focus on the print stream because all of these desktop computers were connected to printers, the ability to capture the PostScript which offered device and platform independence, and finally we needed to make synthetic fonts so we wouldn't have to ship the fonts with the documents. By 1992 we had a prototype that could display on screen fast, but we realized we had to do the file structure exactly right. We used our best computer scientists to build a robust file structure. [Then] we looked around at 20 jillion legacy documents*

on paper and thought it would be great to capture them in this new format. So, in 1993 we bought two OCR companies and re-wrote recognition programs from scratch, mostly using the engineer's expertise from these acquisitions.

The critical invention in the Capture product was that unlike OCR, we didn't use tildes or asterisks to signal uncertain recognition, we put in a bitmap for any possible mistakes. This potentially eliminates the editing required in scanning, and the file is still readable and searchable. Of course, we are currently working to make this a jillion times better. [And] we guarantee that Acrobat will stay cross-platform, stay stable and be a reliable medium for archiving. We are continually adding features to make it a repository for information.

Indexing Documents

When information is scanned or otherwise entered into a DIP system, it is imperative that an effective indexing system is implemented–the data in your system is of little use to anyone if it is not easy to locate. An effective indexing system codes the data with some identifying information so that the stored item can be found in your system's archive using any number of identifying labels or attributes. The type of information that must be recorded about your documents will depend on the ways the documents will be used. If your DIP system will be used primarily for forms processing, perhaps, you will only need to record information such as the form type, form number, account number, and the name of the client/customer to whom the form is relevant, etc. On the other hand, you may need a system that implements full-text searching capabilities in situations in which staff members may need to retrieve all documents related to particular keywords. Before implementing any particular indexing system, it is essential that you identify the various ways that your

documents will be used and the most likely strategies that staff members will use to find the particular documents that they need.

General indexing strategies collect different types of information about documents in various ways. In any indexing system, the location of the document in your DIP system's archive must be recorded. If your system uses various storage devices and integrates various document types, accurately identifying the location of the document will be critical to efficiently retrieving the document from the system. Typically, the initial indexing step will also collect complete bibliographic information for locating and retrieving documents, such as the details listed below:

- Author or source of the document.
- Exact date of capture.
- Location of the capture device.
- Any relevant modification details.
- Appropriate cross-referencing information.
- File type.

In addition, your indexing process will collect structured information–information that is recorded in a standard way for all documents of a particular type. For example, in the case of sales orders, the system might collect order numbers, order dates, customer name, etc. Depending on the sophistication of the scanning/indexing system that you use (and whether or not the system can recognize text in an image), some or all of this information may be collected automatically. On the other hand, these steps may be completed manually by a staff member who manages the scanning of documents. Or your software might allow for some combination of automated and manual input using point-and-click data selection and indexing templates. *FileMagic* is one software product that allows for this type of

automation. The screen below illustrates typical fields used to index a sales order:

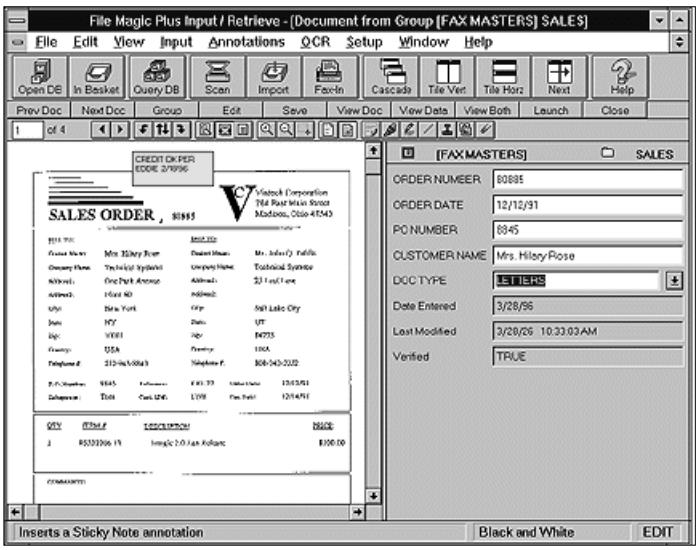


Figure 3.4
File Magic allows for various structured index fields.

Obviously, once this structured information is entered into the index, documents can be located by searching for matches in one or several of the fields. For example, a particular order number can be retrieved from the archive, or all of the sales orders associated with a particular customer, or even all of the orders placed on a certain date.

In addition to this basic indexing, your system may require indexing of various other types and levels of complexity. For example, you may need to index documents based on full-text reviews. In this case, your system must definitely include a conversion utility that creates readable text from images. These text reviews may use extraction indexing methods, whereby documents are indexed according to particular words that appear in the document with some established level of frequency. On the other hand, some organizations need even more sophisticated

indexing. One type of indexing that adds an additional level of complexity is assignment indexing. Assignment indexing includes terms that are not necessarily included in the text, but which are related to words included in the text. If your organization will be using documents in your archive for in-depth research, you may wish to explore these various indexing strategies.

Another indexing need to consider in your DIP system is the possibility of extracting particular information contained within graphics. For example, an organization that includes engineering schematics in its archive will need to index the schematic in terms of the project to which it is related. But, they may also need to index individual components of the image, such as particular symbols. Indexing technologies that extract data from images and use that data for sophisticated indexing and searching are available from various organizations and boast varying levels of accuracy. If your organization plans to use any of these technologies, contact vendors and ask for demos that can help you judge the value of the products in your system. (To learn more about how data is retrieved from raster images, you may also wish to read the informative research paper "Intelligent Data Retrieval from Raster Images of Documents," available at **http://www.csdl.tamu.edu/DL94/paper/srihari.html**.)

Bar Codes

Bar codes were invented in 1962 to allow for the tracking of railroad cars. Since that time, the applications of bar-coding technologies have spread to the automobile industry, grocery stores, package shipments, and even the public library! When indexing research technologies for your DIP system, you may also wish to consider bar codes.

Simply defined, bar codes are patterns of wide and narrow bars, printed on paper or some other material. Computers read bar codes by scanning them (with a laser beam or a wand). Bar codes can be used to encode many types of data, and in high-volume document imaging systems, bar codes are frequently used to automate indexes and document control. For example, when a group of pages constitutes one indexing entity (such as medical claims forms and supporting documents), a bar code of the first page of the group can be used to identify the cluster of pages as related in terms of the DIP indexing scheme. When the first page of the next group of documents (with the affixed bar code) passes through the scanner, the system automatically recognizes a new cluster that requires a new indexing entry.

Obviously, the indexing of data in your system's archive is closely related to the retrieval strategies that your system will use. We will discuss retrieval and its relationship to indexing later in this chapter.

Storing Documents

As we mentioned earlier in this text, one of the primary benefits of implementing a DIP system is the reduced costs of storage space for electronic versions of documents. According to the AICPA publication, *Audit Implications of EDM*,

> *It has been estimated that the cost of owning and maintaining a standard five-drawer file cabinet is $880 annually, with the annual cost-per-filing-inch being $11. At $20 per square foot, such drawers–roughly equivalent to the space of a 16' by 24' room–would cost the entity $450,000 per year to maintain and would occupy several employees. The cost of an optical disk is a fraction of that needed for paper filing, and the system can be run by far fewer people. For example, one financial company had disposed of nearly five tons of paper records and reduced archive space by 60%.*

As the costs of electronic storage medium continue to come down, organizations that considered DIP too costly a few years ago are now re-examining their DIP options. In order for these cost benefits to be successfully realized, accounting professionals working on DIP projects must make careful choices about the storage methods and technologies that will be used in their organizations. Storage devices are typically described in terms of three characteristics: (1) storage capacity, (2) data transfer rate, and (3) access time. Storage capacity identifies how much data can be stored on a device. Data transfer rate identifies how quickly data can be transferred to the storage device. And access time identifies how quickly data can be located on and retrieved from the storage device. Selecting the storage strategies for an organization involves identifying the organization's requirements in each of these three categories and selecting a product that meets the organization's most pressing needs within whatever budget constraints the organization faces. As you might imagine, the costs of storage devices increase with the device's capabilities in each of the areas listed above.

The two main types of devices used for the storage of electronic documents are magnetic and optical. In the sections that follow, we will describe these two types of storage methods.

Magnetic Storage

The most familiar type of data storage tools are magnetic devices. In fact, the hard drive of your computer is a magnetic storage tool, as are the floppy disks you've been using for years! In simplest terms, magnetic storage devices store data by using electromagnets to magnetize particles on the surface of a storage disk. These magnetized particles represent data bits (either 0 or 1) that computers recognize and understand.

Traditional magnetic media, such as conventional hard disk drives, are typically used for short-term storage in image processing systems because they provide fast data access speeds and transfer rates. However, magnetic storage media are not as economically feasible as optical storage media when you are planning for long-time storage. They cannot accommodate the same volumes as optical storage options and they degrade over time.

Even so, high-capacity drives, and technologies such as DAT (Digital Audio Tape) and RAID (Redundant Arrays of Inexpensive Disks) may make magnetic storage a more viable option for some organizations. DAT can hold approximately 1.5 GB and is used in EDM systems primarily as a backup medium and as a master for CD-ROM. Multiple DAT cassette loaders that store approximately 10 GB are also available. A bulk erase utility is employed for tape reuse. RAID is a highly reliable strategy for storing data on a set of disk drives that work together. RAID storage systems may be simple, as in a system where one disk stores a copy of data that is on another disk. On the other hand, some RAID storage systems use sophisticated strategies for storing and protecting data. For example, a RAID system might divide data into bytes (eight bits) and then further divide those bytes into bits, storing each bit on a different disk. Error-correcting code can be stored on separate disks; then, if a single disk in the system fails, only one bit of data per byte will be lost. The error-correcting code can be used to restore the missing bits. (RAID redundancy strategies are also used with optical storage systems.)

Optical Storage

Currently, the most popular storage medium for digital image files is optical disk. Optical disk technology is not new–most of us are familiar with music on CDs, and we are familiar with the CD-ROM drive on our PCs, both of which are common examples of optical disk storage.

Figure 3.5 *CDs are a common example of optical disk storage.*

As we mentioned earlier, creating an image of a document creates a very large file; in fact, a single page can require 50,000 bytes of storage! Fortunately, a typical optical disk can store from 40,000 to 150,000 of these pages. Unlike magnetic storage devices, optical storage devices store information by etching spots into a plastic disk with a focused laser beam.

Most optical disks used in image processing fall into the write-once, read-many (WORM) category. When using WORM, once written to disk, files cannot be erased or changed. Industry experts once thought that the development of erasable optical

media was essential to the acceptance of imaging technologies; however, despite the ready availability of erasable optical media today, WORM storage is far more prevalent. Because of their tremendous storage capacity, and because they can be accessed innumerable times, these systems play an important archiving role. Some 12-inch optical disks can store approximately 20,000 images on each side.

The storage disks themselves can be stored in different ways. Small systems use manually inserted, stand-alone disks and can handle one disk at a time, while larger units use multiple disk drives and can house from 20 to 100 optical disk platters, known as jukeboxes. Just like the musical jukeboxes of the soda shops of the 50s played records, these units use robotics to mount and dismount large numbers of optical disks. A jukebox can house as many as 100 disks and several drives, storing and quickly accessing millions of images. This whole operation often takes only a few seconds. In very large systems, a number of jukeboxes may be interconnected to provide online access to hundreds of millions of images.

Other Storage Options

Often, as businesses migrate from existing storage devices to new components and technologies, a hybrid storage strategy emerges. Some businesses define schedules for retaining records and target storage media for each class of records. One of these hybrids is called magneto-optical storage; this is a rewritable medium that provides the high-density storage of optical media with the revision capabilities of magnetic media. Increasingly, magneto-optical is being implemented in large-scale systems for storing both work in progress and final documents.

While not an imaging system (which requires scanning and indexing), another important technology to consider when discussing image processing strategies is COLD (originally, Computer Output to Laser Disc; now more accurately, Computer Online Data) storage technology. COLD can file, store, and index text-based documents created by a data processing system. So, with this method, you never need to generate paper in the first place. Yet your reports are safely stored away on the COLD computer system for access when you need them. Once limited to laser disk technology, the storage media can be any combination of optical disk, CD, or magnetic disk.

Typically, a COLD system works nicely in conjunction with scanning systems. That is, while your scanning staff converts paper documents into electronic documents, users of a COLD system create electronic documents (such as accounting reports and in-house generated invoices) that are captured and stored electronically. According to Maximal Systems, Inc. (**http://www.maxretriever.com**), a leader in the COLD technology industry, the advantage of a COLD system is clear:

> *COLD documents are generated from your own computer system and captured as data. The entire system is automated, from document capture, to indexing, to storage. You simply benefit from easier access to documents that were previously stored on microfiche, printed on multi-part forms, or distributed on greenbar paper. For example, your organization probably generates invoices or accounting reports [which are typically] stored in file cabinets or on microfiche. Rather than print an in-house copy of invoice forms, or sending a tape out for microfiche, you simply capture a copy of the print file and send it to your PC network.*

Once captured, COLD technology automatically indexes, compresses, and stores the documents on a local or network hard drive. Later, they can be moved to optical, CD-ROM, or DAT storage.

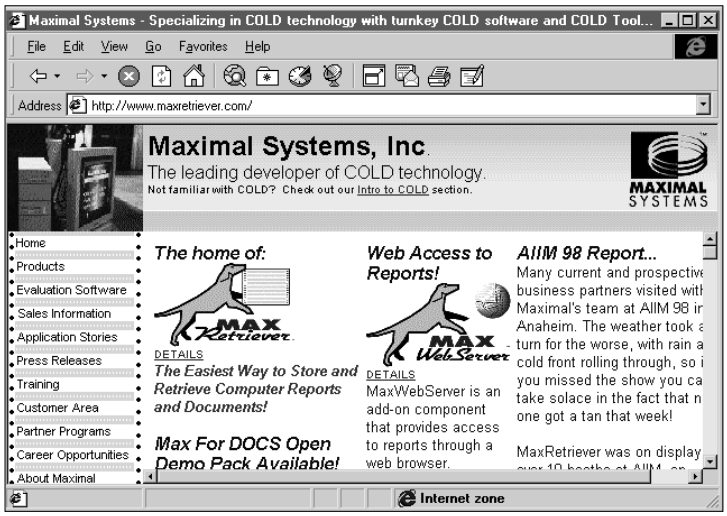


Figure 3.6
Learn more about COLD technology at Maximal Systems, Inc.'s Web site.

Data Compression

An additional concept that you must consider when planning for efficient data storage is compression. Data compression is used to "squeeze" cumbersome files (text, programs, graphics, etc.) into more manageable sizes so that they use less space, reduce download times, or fit onto storage disks.

Data is compressed by software that uses specific algorithms to reduce the sizes of large files, programs, or text documents. The sizes of such files can be compressed in many ways; for example, the compression algorithm may use a single character to replace the "full-stop; space" sequence that typically falls at the end of sentences. The amount of reduction depends on the compression program and the document itself–from most programs you can

usually expect a 50%-87% reduction in document size. If a document contains many graphics, the rate may be lower.

Data is decompressed when a compressed file is expanded to or near its original form by a decompression program. Decompression can, at times, be confusing, because each program uses a unique algorithm to compress files. As a result, you need a corresponding decompression program to restore files. The filename extensions on compressed files indicate which decompression program will restore the file. For example, if a file is compressed with *WinZip*, you must use a program that decompresses *.zip* files to restore the file to its original form.

When planning for compression and decompression on your DIP system, you will need to consider various products and the points at which these processes should take place.

Retrieving Documents

The process of retrieving documents from a DIP system actually consists of two parts: finding the document and producing appropriate output. You may use various searching strategies to locate documents in your system's archive. And, as with other technologies we have overviewed in this chapter, different vendors offer unique interfaces and features in their particular products. Even so, before selecting particular products, it will be helpful to understand document retrieval in a general way.

Typically, when a particular piece of data is needed, a DIP system user accesses a searching utility, enters values that match the desired document (known as a query), and then waits for the searching software to locate and retrieve the document(s) that match the query. Depending on the complexity of the document management system, the search may be limited to a particular area of a network archive, the entire archive, or, in systems with

Internet connections, the search may even be extended to the Internet!

As we mentioned in our earlier discussion of indexing, the ways that documents can be retrieved from your system will depend on the ways that documents have been indexed. For example, if you have only implemented structured indexing, you will most likely have a retrieval system that depends on the fields used in that system. In our earlier example, for instance, we indexed sales orders in terms of items such as order numbers, date, customer name, etc. When locating documents in this system, users must enter values in these preset fields. The system will not recognize any value outside of these fields. You may also need to consider search-and-retrieval tools that allow for refined combinations of fields (and keywords). For example, search engines that support Boolean operators will allow users to limit retrievals to items that match all fields by typing the word "and" between the items (or by selecting this option from a list of choices).

Text retrieval systems retrieve documents based on words and word meanings found within the full content of the document. The retrieval software uses an index created from a database that stores information about the documents in the DIP archive. To increase access speeds, the document index database is often stored on magnetic disk, apart from the documents. The index database may also be stored on a separate dedicated server or on a host mainframe, apart from the DIP system.

There are various text-retrieval architectures; the three most common, stand-alone, host-based, and networked, offer vast differences in pricing and functionality. The most common text retrieval system, the stand alone, is a PC-based architecture with relatively inexpensive components. Stand-alone systems offer a wide range of features and capabilities; however, there are also

some potential problem areas. For example, often searches are limited to a single document collection and low-end products are typically integrated with an interface that cannot be altered freely. Host-based systems allow multiple users to simultaneously access the document collection and query the system, but all of the documents must reside on a central platform. All processing and index creation is performed on the host machine; and the end-user platform functions as a dumb terminal. In a networked retrieval system, users access a single version of the document collection located on one node of the network and use the text retrieval engine on that same node to perform the search. Results are downloaded to the local node, where viewing and editing operations can be performed.

The type of search-and-retrieval system used in your DIP system will depend on the types of documents included in your archive and the ways staff members will need to use that data. Be sure to take the time to assess your organization's needs before selecting any particular system. Many document management products that include integrated searching utilities offer demonstrations that can help you. For example, the screen shown below illustrates the searching utility of OpenText's *Livelink Intranet*. A guided tour of this product can be viewed at OpenText's Web site (**http://www.opentext.com**).

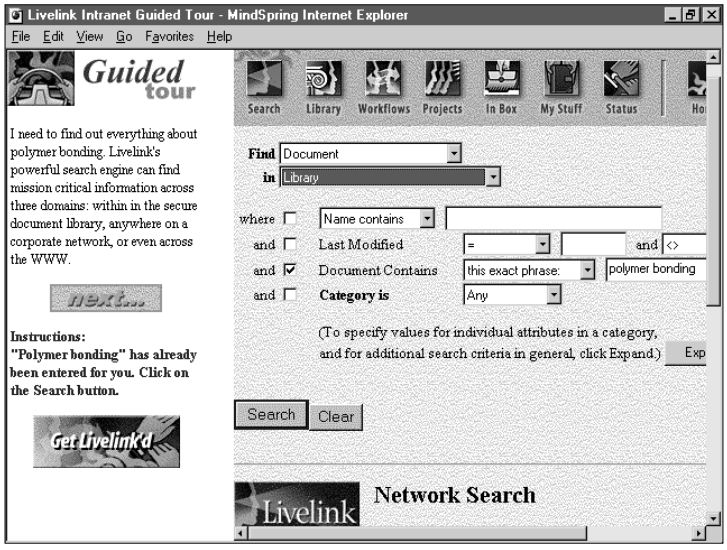


Figure 3.7 *Livelink allows network and Internet searches.*

You may even wish to distribute a questionnaire to staff members to identify the ways that they use documents when examining your various options.

Output

After a document has been located, the data must be output in some way. Data output options include viewing, printing, and transmitting electronic files. Once the desired file is located, it can be viewed at a display station (typically a PC monitor), printed, sent to a fax gateway or even to an e-mail program for distribution. As we discussed earlier, the ways that data will be output from the documents retrieved from your archive will have determined the settings used to capture and create those documents. If you are developing a DIP system built on existing hardware with output limitations, the documents in your archive will have been created to work within the limitations of the system. On the other hand, you may need to consider upgrades and hardware purchases to make the most efficient possible use of output on your system. For example, complex images, such as

engineering schematics or imaging used in medical research, require sophisticated monitors to display and PCs with enhanced memory, video cards, and graphics-caching capabilities. Otherwise, the efficiency that you gain by making these documents available will be undermined by the efficiency lost as staff members wait for images to display on workstations that are not equipped to handle them.

Network Considerations

Although it is beyond the scope of this text to discuss the networks on which DIP systems are built, it is important for DIP project managers to recognize that DIP systems can place great strain on existing communications systems. The fact is that digital images, even when they have been compressed, are quite large. If you are constructing a DIP system on an organization's existing network backbone, a network backbone not intended to manage the transmission of large volumes of data, system response times will be disappointing. Be certain to evaluate the strengths and limitations of the existing network backbone and make sufficient upgrades to get the most out of your DIP system. (In Chapter Six, we will discuss guidelines for performing a site analysis.)

Assessing DIP Requirements

The worksheet on the following page identifies the major components of DIP systems. In the boxes to the right of each technology, identify your organization's current methods for completing the tasks that fall into each category and the particular requirements of your organization. For example, your current practice for data capture might be manual data entry. The needs of your organization might be increased speed and accuracy. Indexing may be handled through a manual hard-copy filing system. Your needs in this area might include easier accessibility and more detailed indexing capabilities. Considering these needs now will help you make appropriate hardware and software choices later.

Identifying DIP Requirements

DIP Component	Current Practice	Current (and Future) Needs
Data Capture		
Data Indexing		
Data Storage		
Data Retrieval		

Executive Summary

In this chapter, we have overviewed the four major components of Document Image Processing systems (capturing, indexing, storing, and retrieving). In addition, we have described the major technologies that are used in each of these four system components. Understanding these concepts in a general way will prepare CPAs to make sensible choices about system development strategies that may be implemented in DIP projects. And understanding the technologies that are driving the development of related hardware and software products will help CPAs wade through vendor hype and examine the important features of DIP products. Finally, by examining the components of a DIP system, we have drawn attention to the various levels of decisions that must be made before any DIP technology can be efficiently implemented.

Chapter Four
Selecting DIP Tools

Once you decide to implement a DIP system, you will need to determine which hardware and software components best suit your working environment. As with other technologies, there are numerous options from which to select, and there isn't necessarily a universal best answer. In Chapter Three, you learned how DIP components function; in this chapter, we'll overview some of the particular products that support various phases of a DIP system. Obviously, comprehensive coverage of every vendor and product is impossible; even so, by examining a few products from selected product categories, you will come to understand how hardware and software support the DIP process and features that you might expect from the components you choose. In short, although your research will lead you to vendors and products not discussed in this chapter, overviewing even a limited number of products and vendors will help you to make sensible selections for your organization. Topics covered in this chapter include the following:

- Integrated software suites.
- Scanners.
- Imaging support software.
- Storage tools.
- Product selection criteria.
- Other vendor resources.

Integrated DIP Suites

In our overview of DIP in Chapter Three, we broke the entire process into four phases: capturing data, indexing data, storing data, and retrieving data. Although these categories are very useful when your goal is to understand DIP in a general way, the products that support the DIP process are not as neatly segregated. In fact, it is unlikely that you will purchase products that support data capture without considering data indexing, storage, and retrieval. You may be looking for DIP products that can build on existing hardware and software; and you must select DIP products that are compatible with your network platform. Then, in order for a DIP system to function efficiently, you will need to consider the entire process and strive for integration and compatibility between various process phases.

If your organization performs very focused and specific types of tasks, you may be shopping for DIP functionality in new versions of your organization's primary software tools. For example, many of the new high-end, client/server, accounting packages include document management and document imaging components as part of their suites of functions. In the March 1998, *CPA Technology Advisor* (**http://www.cpatech.hbpp.com**), George Wilson, Jr., CPA, identifies a number of accounting software packages that include DIP functionality. We have listed these products, and the Web sites where more information is available, below:

- Computron Software's *Computron Software* (**http://www.computronsoftware.com**).
- FlexiInternational's *FlexiFinancials* (**http://www.flexi.com**).

- Geac SmartStream's *SmartStream* (**http://www.smartstream.geac.com**).
- JBA International's *System 21 Financial* (**http://www.jbaintl.com**).
- J.D. Edwards' OneWorld (**http://www.jdedwards.com**).
- Oracle's *Oracle Applications* (**http://www.oracle.com**).
- Peoplesoft's *Peoplesoft* (**http://www.peopesoft.com**).
- Prestige Software International's *Masterpiece/Net* (**http://www.ca-masterpiece.com**).
- Quality Software Products' *QSP Financials* (**http://www.qspinc.com**).
- Ross Systems' *RenaissanceCS Financials* (**http://www.rossinc.com**).
- SAP America's *R/3* (**http://www.sap.com**).
- Scala North America's *Scala* (**http://www.scala-na.com**).
- Walker's *Tamaris* (**http://www.walker.com**).

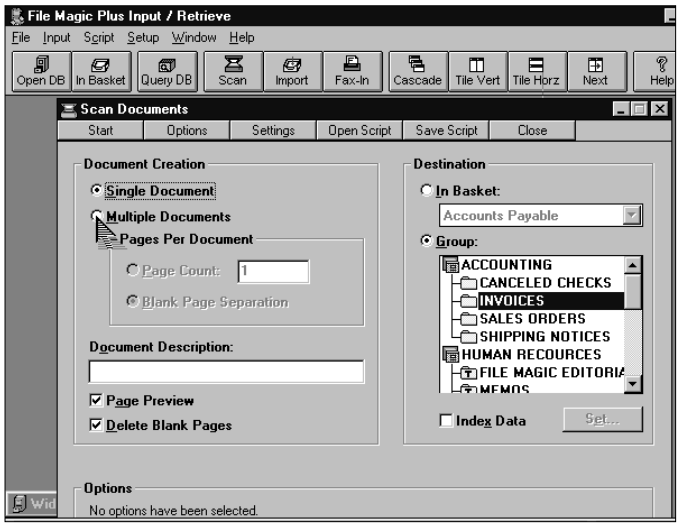
If your needs are more broad, and you are seeking an overall integrated DIP/EDM system, many software vendors have made this easy by offering integrated software suites to support the entire DIP process. In some cases, these suites combine separate units of a "family" of products that may also be purchased individually as your system grows. In the documentation for these products, you will find listings of hardware components with which the suites are compatible. Be certain to check these lists before purchasing new hardware to verify that your existing hardware will function adequately with the new suites. In the sections that follow, we will overview the features and system requirements of several of the better-known DIP/EDM suites.

Westbrook Technologies' File Magic

Document management industry leader Westbrook Technologies offers a solid family of DIP and EDM products for organizations of various sizes (**http://www.filemagic.com**). Among their

integrated products *File Magic Plus* stands out as a useful example of a suite that blends and supports all aspects of a DIP process. With this system, you can capture, index, file, retrieve, and edit and annotate information from within one *Windows*-based software interface. For example, the screen below illustrates one of the *File Magic Plus* screens that controls the capture of data through the scanning process:

Figure 4.1
Capturing data in File Magic Plus.



A similar interface guides users through data imports from desktop applications (such as spreadsheet and word-processing programs), fax, and even the Internet. Indexing steps are just as easy to complete through the *File Magic Plus* interface. The screen below illustrates a sample document and the indexing fields used to begin the indexing process:

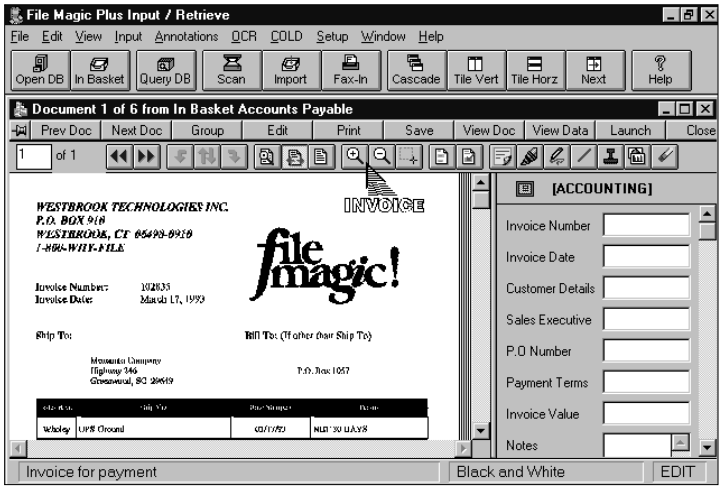


Figure 4.2 *Supplying indexing information in File Magic Plus.*

The indexing fields can be automatically or manually completed using various time-saving features, such as point-and-shoot indexing. This *File Magic Plus* feature allows a user to identify an image area and then "shoot" the data in that image area into an indexing field. Once a document has been indexed and stored, *File Magic Plus* also offers a retrieval tool as part of the suite. The screen below illustrates a search that allows the user to locate an archived document by entering criteria from three indexing fields:

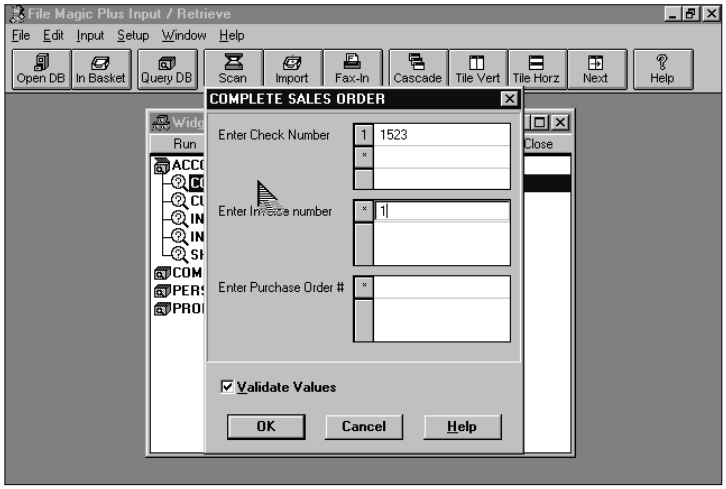


Figure 4.3 *Performing a search through File Magic Plus.*

Once a user locates a file and retrieves it from storage, the file can be annotated in any number of ways using the *File Magic Plus* editing toolbar. As in other *Windows* interfaces, to use an editing tool, simply click on the button that represents the tool and use your mouse to complete the editing task. The screen below shows the *File Magic Plus* editing toolbar and a document that has been annotated with notes and highlighting during the DIP process:

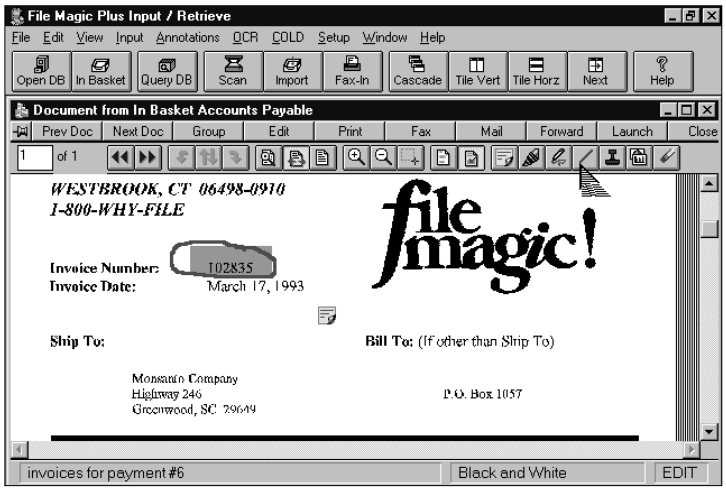


Figure 4.4
Annotating a document in File Magic Plus.

Minimum requirements for running *File Magic Plus* include an IBM compatible PC with a 486 processor and 8 MB of RAM. The client platform must be *Windows 3.1* or higher. With the appropriate user-count licensing, system administrators have the additional capabilities of creating user accounts and in-baskets (either public or restricted), and a variety of other system security functions. For enterprise users, the *File Magic Fortis* edition may be the answer. The interface is similar and integrates easily with both *Windows NT* and *Novell 3.x-4.x* network platforms.

As mentioned earlier, suites of products often include various add-ons that can enhance the base system. For example, users of *File Magic Plus* who must also create CDs for distribution or portable storage can use the *CDExpress* product. This product allows the same DIP functions described above along with features and tools that allow for database information to be retrieved, annotated, and "burned" to CD disks for portability, distribution, and storage. The diagram below illustrates the various components of *CDExpress* and the flow of data that the system supports.

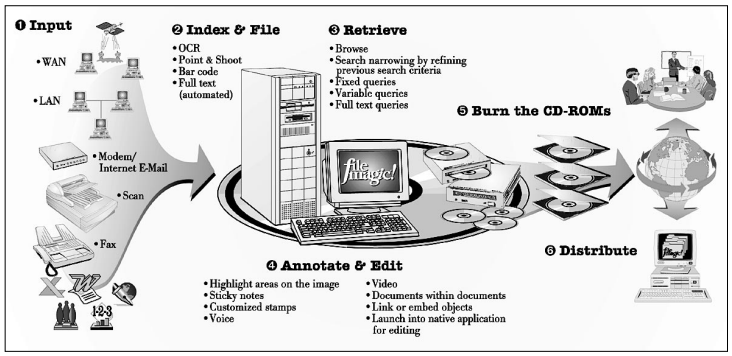


Figure 4.5
File Magic's CDExpress.

As the above illustration shows, *CDExpress* functions on a six-step process, integrating the various components of the DIP system into a complete document management system:

- You can input information from any number of original forms; scanned documents, fax, WAN, LAN, or word-processing and spreadsheet programs.
- Once entered, documents are indexed and filed in the method of choice; OCR, or full-text automated, for example.
- Retrieval tools, such as browsers and fixed queries allow you to quickly locate any document in the system.

- The editing tools allow you to add video, sticky notes, and voice; as well as editing from your own word-processing program.
- Record the CD-ROMs from the desired database or index.
- The CD-ROM can be stored in-house (in an optical jukebox, for instance) or you can distribute the CD-ROM in any manner you wish.

To learn more about these and other Westbrook Technologies, Inc., products, visit their Web site at **http://www.filemagic.com**.

FileNET's Panagon IDM Desktop

Another example of an "out of the box" software package, is FileNET's *Panagon IDM Desktop* suite of applications. This suite includes various utilities that allow users to control DIP and EDM functions through a versatile integrated interface. For example, the *Panagon Capture* component can be used to control the capture of a wide range of document types–scanned paper, fax, e-mail, HTML, audio, word-processed, etc. Once entered into the system, *Panagon IDM Desktop* simplifies access, management, and control of all document types throughout the system; and this program is fully functional in both client/server and Internet environments. *Panagon IDM Desktop* has an intuitive interface that allows you to access any document via the "FileNET Neighborhood" on your desktop–*Windows* users will be familiar with this hierarchy.

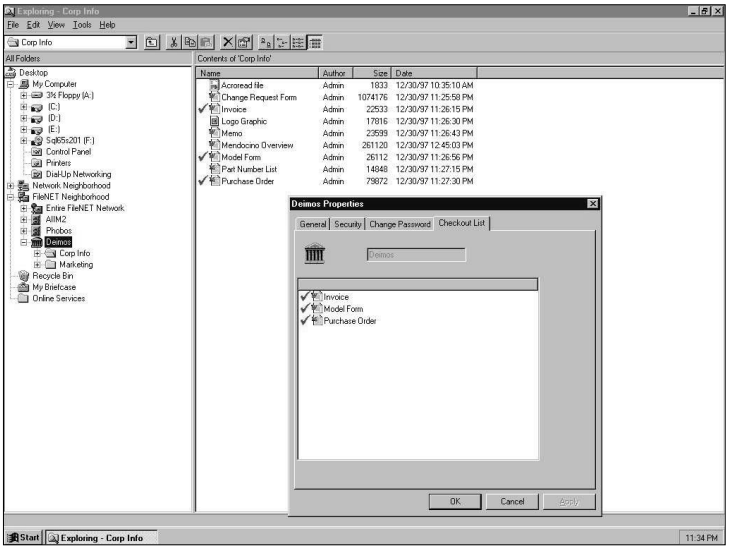


Figure 4.6 *Exploring the FileNET Neighborhood.*

Files can simply be selected from the *Explorer* hierarchy and displayed in the *Panagon* desktop, as shown below:

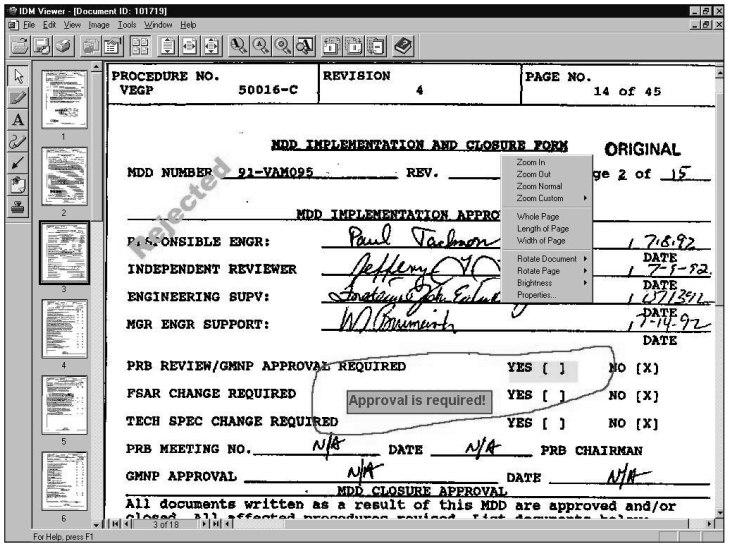


Figure 4.7 *Displaying a scanned document in the Panagon IDM Desktop.*

Panagon can display more than 200 data types, even when the original applications are not running on the client machine. This familiar *Windows* interface offers the obvious advantage of reduced training time for system users and the added comfort of a familiar computing environment. If your staff is more comfortable with a Web interface, *Panagon IDM Desktop* can also display through either *Internet Explorer* or *Netscape Navigator* windows.

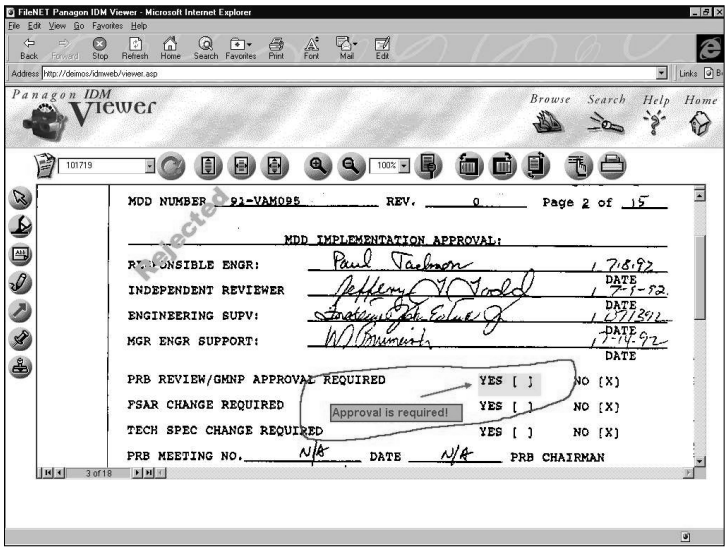


Figure 4.8 *Viewing a scanned document through a Panagon/Web interface.*

FileNET and FileNET clients make much of this capability. In fact, FileNET boasts "dual deployment capability [that] is suited for 'thin' clients connected via the Internet and traditional 'thick' client/server environments. This means that no matter where you are, you will always be able to get to your information, confident that it is accurate, up-to-date, and working for you." (In Chapter Five, we will look at other Web-enhanced DIP options.)

As with *File Magic Plus*, the FileNET family of products also offers integrated storage solutions that you may wish to consider when exploring this product suite. For example, *Panagon Report*

Manager is a client/server COLD product that indexes, stores, retrieves, prints, faxes, and distributes computer-generated output on magnetic or optical disks.

System requirements for desktop machines running *Panagon* client software include a Pentium processor, high resolution display, 50 MB of free disk space, and 32 MB of RAM. The platform can be either *Windows 95* or *Windows NT 4.0*. If you will be using the Web browser interface, you must have either *Internet Explorer 3.0 or Netscape Navigator 3.0*. The server for the system must be running *FileNET Mezzanine 4.1.1* or greater, *FileNET IMS 3.3.1* or greater, or *Microsoft's Internet Information Server 3.0* or greater. To learn more about FileNET products, visit their Web site at **http://www.filenet.com**. (The screen shots in this section can be found at FileNET's Web site and are reproduced here by permission of FileNET. FileNET is a registered trademark and *Panagon* is a trademark of FileNET Corporation.)

PC DOCS' DOCS Open

The last well-known vendor we will look at in this overview of integrated products is PC DOCS' (**http://www.pcdocs.com**). The PC DOCS flagship product, *DOCS Open,* is another powerful, integrated document management suite. This program allows users to control, organize, access, and share vital corporate information. *DOCS Open* boasts an extended enterprise architecture that allows any number of add-on programs–as your document management needs evolve, so will *DOCS Open*. PC DOCS is particularly proud of the program's retrieval tools:

> *A Document Profile is associated with every DOCS Open document that you create. The Document Profile form is analogous to a library catalog card. It contains customizable attribute information–author, document title, project, application type, etc.–which describes a DOCS Open document. When you want to retrieve a*

document, you simply select a search that locates the Document Profiles that were created by you or other DOCS Open users.

The screen below shows a Document Profile form for a *Word* document:

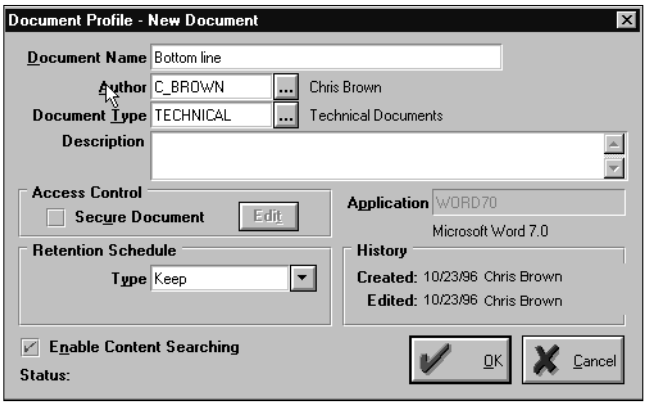


Figure 4.9 *Creating a Document Profile in DOCS Open.*

Another feature that enhances this product's search-and-retrieval capabilities is Quick Retrieve. This feature automatically identifies and lists the last 30 documents edited on the workstation.

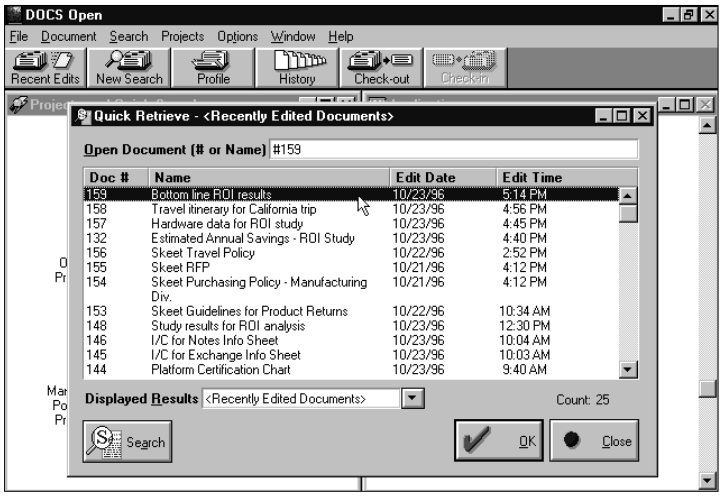


Figure 4.10 *Using Quick Retrieve to view the most recently used documents.*

Other features of the *DOCS Open* product that may add to its appeal for your organization are listed and described briefly below:

- **Attribute Indexing**—This feature allows users to locate documents matching any entries in any one of the fields on the Document Profile.
- **Content Search**—This feature allows users to store frequently used searches for read access. This feature is easily accessed from *DOCS Open*'s Desktop Search menu or through the icon on the desktop.
- **Global Network Searching**—This feature allows users to search on WANs and LANs with equal ease.
- **Version Tracking**—This feature allows users to store as many as 90 major versions of a document, as well as up to 26 sub-versions of each major version.
- **Form Selection**—This feature allows an administrator to set up multiple Profile Entry and Profile Search forms.
- **Security**—This feature provides managers to control who accesses documents and folders and when, and how those documents and folders are accessed.
- **Check-in/Check-out**—This feature tracks documents that users have copied from the network, and it notes when those documents will be returned. Additionally, this feature allows managers to lock specific documents when no further editing or annotating is wanted.
- **Document History**—This feature provides a log of every activity performed on a document.
- **Projects**—This feature allows users to group related documents logically in folders.
- **Templates**—This feature allows users to save forms or boilerplate text for easy retrieval and use.

- **Storage Management**—This feature enables users to delete or archive documents after a specified period of time.

The imaging capabilities of *DOCS Open* are acquired through an add-on product known as *DOCS Imaging*. As you might imagine, this product simply builds document capture capabilities into the *DOCS Open* system. The screen below illustrates a multi-page scanned document displayed in the *DOCS Open* environment through the *DOCS Imaging* add-on utility.

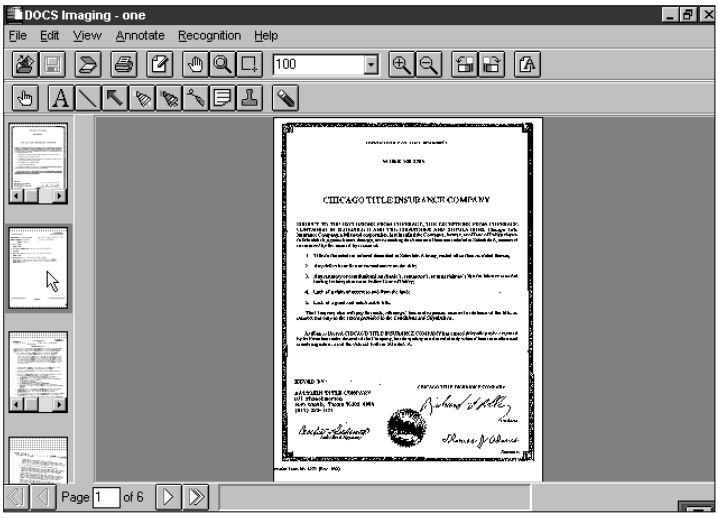


Figure 4.11 *Viewing a scanned image in DOCS Imaging.*

DOCS Open offers an open architecture that supports all leading networks, databases, and desktop operating systems, as well as multiple test, imaging, and workflow engines. Minimum requirements for a *DOCS Imaging* desktop include an Intel processor or compatible system of 486/66 or higher, 16 MB of RAM, and 20 MB of hard disk space. To learn more about PC DOCS products, visit their Web site at **http://www.pcdocs.com**.

Assessing Other System Tools

In many cases, the software suite that you select will lead you to select various types of hardware. As mentioned earlier, software suites may be compatible with only certain types of hardware. As a result, you must make yourself aware of these restrictions when you select a software suite. On the other hand, even within these restrictions, you will find yourself facing many choices about hardware and other software options. In the sections that follow, we will overview some of the criteria that you should consider when selecting other DIP support technologies. As before, our goal is neither comprehensive coverage nor product recommendations. Rather, we will look at a few specific products as examples of the types of features that may make a product more or less valuable, and more or less expensive.

Scanners

As you remember from Chapter Three, there are many ways to capture digital data. Depending on the needs of your organization, you may need to consider various types of hardware. For example, an insurance company that handles car accident claims may wish to explore digital cameras that will allow agents to photograph accident scenes and damaged vehicles and then upload the digital photo directly to the DIP system. As you might imagine, many vendors are currently working hard to develop affordable digital cameras that offer good quality digital photos, such as Canon (**http://www.canon.com**) and Kodak (**http://www.kodak.com**). Even so, many organizations are finding that it is more cost efficient to scan photographs. In fact, the scanner may be one of the most important hardware components that you purchase for your DIP system. As a result, it is important to approach this purchase well informed and with clear criteria for selection.

Obviously, the type of scanner(s) that you choose will depend on how you plan to use it. Flatbed scanners, which look something like photo copying machines, can handle both individual sheets and images in books. Sheet-fed scanners are smaller and may fit better in a cramped home office, but most can handle only unbound documents. Photo scanners are small sheet-fed scanners limited to scanning photos that are, at most, four to five inches wide. Hand-held scanners are the least likely choice for most business environments. So the most likely scanner option for many companies implementing a high-volume DIP system will be a flatbed scanner.

Vendors frequently offer "package" deals–that is, scanners that come equipped with the necessary software to get your DIP system up and running with a minimum of fuss. For many small and mid-size companies, this is an ideal arrangement. And, often you can obtain the information you need online. For example, the Umax Web site (**http://www.umax.com**) provides a wealth of information on scanning products and services for nearly any size business.

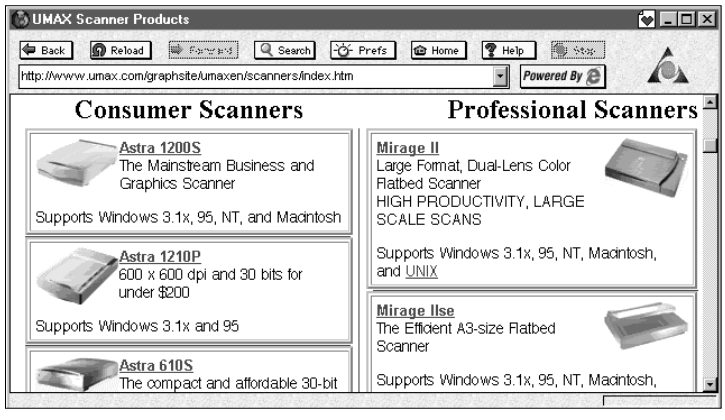


Figure 4.12
The Umax Web site.

An example of an affordable mid-level, color, flatbed scanner that could be a good choice for a small firm or office is the Umax Astra 6005. This modestly priced scanner provides the following features:

- SCSI-2 interface, which makes the scanner compatible with both Macintosh and PC computers.
- Single-pass scanning, which offers accurate color registration in a single efficient scan.
- 30-bit color, which results in images with rich detail in the highlights and shadows.
- 4800 maximum dpi, which provides high resolution for smooth line art and image magnification.
- *VistaScan* software, which offers easy control of scan settings through a *Windows* environment.
- Enclosed optical system, which protects images from distortions caused by dust and electronic noise.
- Cold cathode lamp, which reduces maintenance costs due to long bulb life.
- Additional bundled software, including *Adobe PhotoDeluxe* and *Presto!PageManager*.

Like most scanners of its ilk, the Astra 6005 is simple to set up; just flip a switch located near the rear to unlock the carriage, install the provided SCSI card in any 16-bit slot, and attach the SCSI and power cables–and you're ready to scan! The bundled software includes NewSoft's *Presto!PageManager* program, which allows you to acquire documents and images, and maintain them as electronic files.

No matter which platform of computer you are using, this scanner will give you the power to scan forms and documents for electronic storage, and to scan images for desktop publishing or

Web sites. And the bundled software included with this scanner ensures a quick and easy startup of your DIP system.

Another well-known developer of scanners is Xerox (**http://www.xerox.com**). From their Web site, potential customers can select from scanner models developed for organizations of various sizes. For example, the Document Centre System scanners are designed to support the printing, scanning, faxing, and copying of documents for workgroups of up to 50 people. On the other hand, the Xerox DocuImage 620S boasts fast (up to 40 one-sided pages a minute at 200 dpi), powerful, intelligent image processing capabilities. With possible resolutions as high as 600 dpi, the scanner can contribute to the accuracy of OCR conversions, if your DIP system will use OCR/ICR technologies as part of the processing of documents. The scanner can be automatically fed standard-sized documents, or a by-pass feeder can be used for odd-sized documents. To learn more about the Xerox family of scanners, visit the *Xerox Scanning Solutions* Web site at **http://www.xerox.com/scan/index.html**.

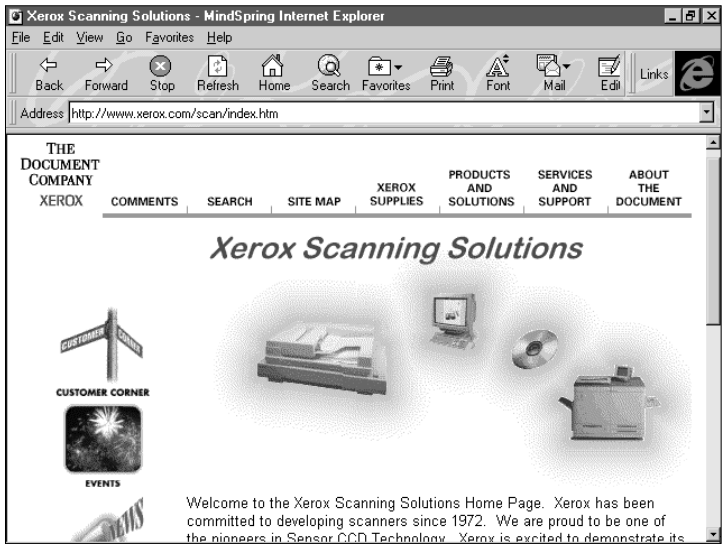


Figure 4.13 *Xerox Scanning Solutions Web site.*

Hewlett Packard (**http://www.hp.com**) offers scanners of various types and sizes, including a well-reviewed scanner for networks of 50 or more users. For example, the HP Network ScanJet 5 has been favorably reviewed by publications such as *Computer Shopper*. ScanJet 5 supports multiple-page scanning, optical character recognition, and outgoing faxing when used in conjunction with a network fax gateway. Depending on the resolution required by your system, ScanJet can scan 15 pages per minute. ScanJet is compatible with *Windows NT* and *Novell Netware* networks and offers a 25-user license in the purchase price. To learn more about this and other Hewlett Packard scanners, visit the Hewlett Packard Web site.

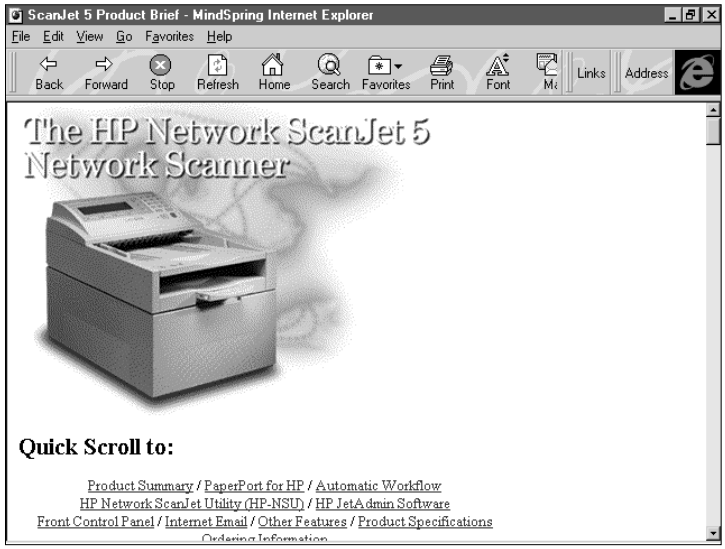


Figure 4.14 *Researching scanners at the Hewlett Packard site.*

Selection Criteria

Like any equipment investment, shopping for the best product at the best price can be frustrating and time consuming. However, there are several ways to make the experience less troublesome. For example, Sandi Smith, CPA, CMA, CDP, recommends asking the following questions as you begin to shop for a scanner:

- How many documents can the scanner process per minute?
- What size paper will it scan?
- What is the scanning resolution?
- How large are the image files? Is a compression routine used?
- For high-volume operations, are there auto-feed and dual-sided processing capabilities?
- Can color images be processed?

And, as you narrow your search for a scanner, there are some (often) less obvious questions you should pose. According to an industry white paper by Michael Bida, Product Manager for Scanners and Recognition Products at Eastman Kodak's Business Imaging Systems division, "hidden factors associated with the scanning process are as important to the bottom line as the more obvious ones." The white paper, "Estimating the Cost of Scanning: It May Not Be As Easy As You Think," lists the following six areas that are frequently overlooked by those shopping for scanners:

1. Labor costs—Records managers with several capture devices may be able to dramatically cut costs by consolidating equipment and upgrading to mid- or high-volume scanners. Since each scanner requires a dedicated operator, labor costs rise or fall in direct proportion to the number of scanners in the system. And managers should remember that it is not the operation's average volume, but rather its peak volume that weighs heaviest when selecting a scanner.
2. Reliability affects throughput—Managers should ask vendors to provide mean-time-between-failure figures. Another factor that is often overlooked is ease of maintenance; that is, every scanner requires routine

maintenance (paper dust must be removed and rollers cleaned at regular intervals); if the primary parts of a scanner are difficult to access, operators tend to avoid cleaning the machine. Obviously, a poorly maintained machine will not perform at its maximum efficiency–and results in more frequent service calls.

3. Service plans—Managers should determine whether service is provided by a manufacturer or distributor, what kinds of service programs are offered, and what level of service is required for a scanner. Records managers who do not work through an integrator should find a service provider that is equipped to address software and hardware problems.
4. Throughput—While most vendors determine their throughput rate by measuring the raw transport speed, this figure does not include gaps between documents, type of document scanned and sizes of batches, etc. An accurate throughput estimate is best determined by having a vendor, or integrator, perform a benchmark test using a recent day's worth of work (preferably a peak day).
5. Ergonomics—Scanners should be designed so that operators can perform as many tasks as possible from a sitting position. And, controls should be located for easy access; and operators should be able to view documents being scanned (and retrieve documents after scanning) while seated at their workstations.
6. Integration—The type of interface a scanner features has a direct impact on integration costs. A SCSI interface facilitates cost-effective integration; while a scanner that requires a video board can dramatically increase the cost of a system. For example, a video board costs approximately $5,000–and a duplex scanner requires two boards.

As well as the items listed above, Bida urges caution when using recycled paper in many scanners, because "while good for the environment, [it] causes havoc with many scanners [because] it contains a binding chemical that damages rollers and can cause jams." If your company is a dedicated user of recycled paper, then chose a manufacturer that has "designed scanner rollers and cleaning fluids to accommodate recycled paper. It will lead to increased reliability and throughput in the future."

You can read Bida's document in its entirety by accessing **http://www.kodak.com**.

Imaging Software

When you have found the right scanner for your company's needs, you will need the appropriate software to complete your DIP tasks. Typically, your scanner will require three types of software:

- Driver software that controls the scanner's operation.
- OCR software that reads images of text pages and converts text into an editable word-processed file.
- Image editing software that lets you manipulate an image after the final scan.

Most often, the required software will be part of the scanner "package" at the time of purchase. Or these capabilities will be part of the DIP software suite you have selected. On the other hand, if you are working with a small organization, building DIP functionality a little at a time, you may need to consider these various tools separately. Understanding the functions of these tools when they are considered as free-standing units may also help you evaluate the utilities that perform these functions in a software suite that you are considering. Fortunately, manufacturers and vendors are eager for your business and

provide reams of information to help your make your DIP software decisions.

Driver Software

Simply put, driver software applications extend the computer's operating system in order to support a specific device–in this case a scanner. Unless you replace the scanner itself, you probably will not need to make a further investment in a driver device.

OCR (Optical Character Recognition) Software

Importing paper documents into a database requires a highly accurate system. Like other scanner software, often OCR software is included with flatbed and sheet-fed scanners–typically these are provided as bundled software packages, or sold as suites of programs.

As you will remember from Chapter Three, OCR programs convert scanned documents into editable and searchable computer files. Most OCR programs work from within popular applications such as Microsoft *Word* or Corel's *WordPerfect* word-processing programs, but other OCR output options are also under development. For example, according to a recent Xerox press release (**http://www.xerox.com**), Xerox and Adobe Systems have recently combined efforts to produce advanced OCR capabilities.

Many companies are looking for fast and efficient ways to turn paper documents into electronic documents. Scanning them and saving them as TIFF files is little more than a form of electronic microfilm. Scanning them and converting them to text documents with standard optical character recognition packages is a workable solution for some situations. However, for those companies and businesses that want to standardize on the PDF format, Adobe's Capture offers a [different] solution. Unlike classic OCR software

packages, Adobe's Capture software turns scanned images (TIFFs) into PDF files. The resultant files are viewable with the Adobe Acrobat reader. The software can convert text in the scanned image into the appropriate typeface, passing through graphics in the file untouched. Once the file is converted, it will look like the original scan.

As you might imagine, manufacturers and vendors are currently in fierce competition for shares of this burgeoning market. Consequently, you can find as many software applications as you need. And, as your DIP needs increase, so will your OCR needs; so industry leaders are in a fever to provide bigger and better packages and suites of programs that will allow for future upgrades.

For example, Xerox Corporation's software division, ScanSoft, Inc., hails their *TextBridge Pro98* as the software that will successfully allow you to upgrade any scanner's OCR capabilities. According to Xerox's online promotional material (**http://www.xerox.com**), *TextBridge Pro98* boasts myriad features, including the following:

- Advanced zone editing to provide more control over desired results with options for creating, reshaping, resizing, and renumbering zones.
- Improved batch processing that provides the ability to automatically process multiple jobs.
- TextBridge OCR Wizard to walk the new or infrequent OCR user through the process.
- *Adobe Acrobat* PDF support.
- Improved character accuracy.
- Improved document recognition for complex document elements like insets, line art, un-ruled tables, reversed text, and drop caps.

Additionally, this software vendor provides unlimited, toll-free technical assistance–a plus when those inevitable questions arise.

Another leading developer of OCR software, Ligature Ltd. (**http://www.ligatureltd.com**), recently announced a product called *OCR-On-A-Chip*. According to the manufacturer, this product "provides reading capabilities to any machine or integrated system without the need for a PC or extensive memory. [With *OCR-On-A Chip*] OCR will now be available for use in industries such as robotics, digital camera, heavy industry, manufacturing, and consumer product development."

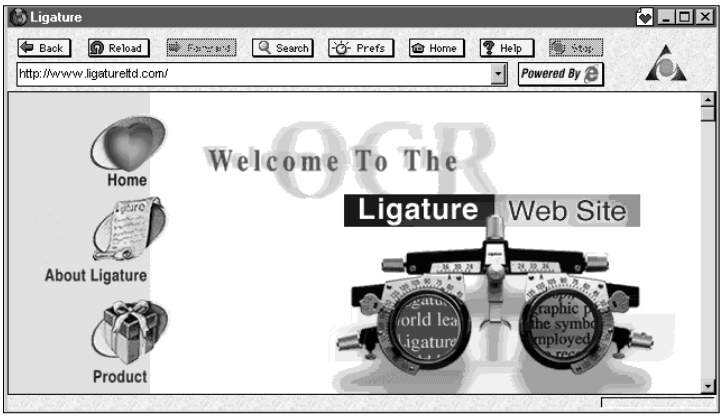


Figure 4.15 *Ligature Ltd. Web site.*

Image Editing Software

One of the most important components of a DIP system is the image editing software. For example, this software must allow the user to select the type of image to create and control the image resolution. Indeed, industry analysts agree that the most important issues for image editing software users are accuracy and ease-of-use. And, like driver and OCR software, your scanner will probably include an image-editing tool. Often, vendors will work with a company to create a customized combination of hardware

and software–thus the consumer can take advantage of "out of the box" document management tools.

For example, the Umax scanner we discussed earlier in this chapter is equipped with such a software "package;" NewSoft's image editing tool, *Presto!PageManager*, is an excellent example of an application that allows you to maintain your files in sets of folders that you can organize as you wish. The left side of the *PageManager* screen displays icons of the images contained within the folders. When double-clicked, an icon loads the document into the main display screen. Once loaded, the document can be altered by using the program's set of editing tools.

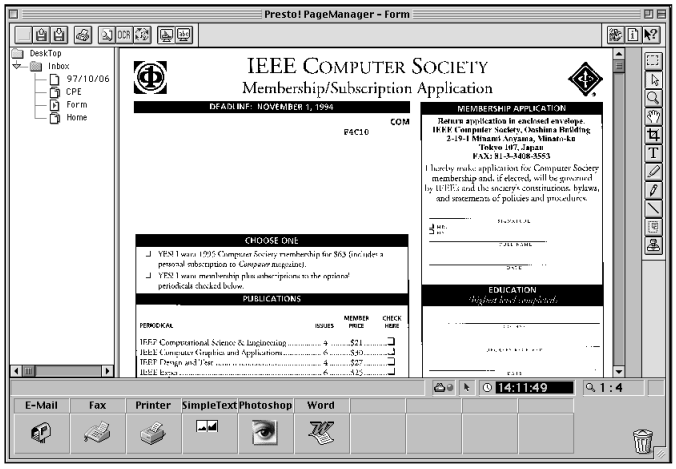


Figure 4.16 *Editing a scanned document in Presto! PageManager.*

The editing toolbar contains a variety of tools that make editing your image a simple process. The functions performed by these tools include the following:

- **Image Select**—Selects a portion of the image for altering.
- **Select**—Selects a particular object for manipulation.
- **Zoom**—Enlarges/reduces the image size on the screen.

- **Move**—Moves images using the drag-and-drop method.
- **Crop**—Defines a cropping frame.
- **Text**—Allows you to enter and edit text.
- **Highlighter**—Highlights a selected area.
- **Freehand Line**—Allows you to draw a "freehand" line on the image.
- **Straight Line**—Allows you to draw a straight line on the image.
- **Sticky Note**—Inserts a note over the image.
- **Stamp**—Applies preset text to the image.

NewSoft, Inc. (**http://www.newsoftinc.com**) is one of the industry leaders in scanning, image editing, OCR, and related softwares. Like many of these online vendors, NewSoft, Inc. offers trial versions of many of their products. Visit their Web site to learn more and to test various products that may benefit your organization.

Figure 4.17
NewSoft, Inc.'s Web site.

(Note: Some software manufacturers like Xerox have incorporated OCR and image-editing tools for greater ease. Xerox's TextBridge Pro 98, for example, allows users to edit their scanned documents with their existing word processor, spreadsheet or

HTML editor. This advanced functionality is not found in most bundled OCR packages included with many scanners.)

Considering Storage Options

So you've researched software suites, scanners and their related software components; now you must also consider the various storage options that will be available to you. As mentioned earlier, in part, these choices may arise naturally from the add-on systems that are available with the software suite you have selected. Or you may be making this decision independently. However you proceed, you will want to review some of the options offered by major vendors before making a final decision.

When exploring storage options, whether optical or magneto-optical, another vender to consider is Hewlett Packard. For example, from the *Information Storage* page shown below (**http://www.hp.com/storgae/optical/main.html**), you can link to pages that provide more in-depth descriptions of the capabilities and requirements of the entire family of HP optical storage options as well as helpful decision-making tools.

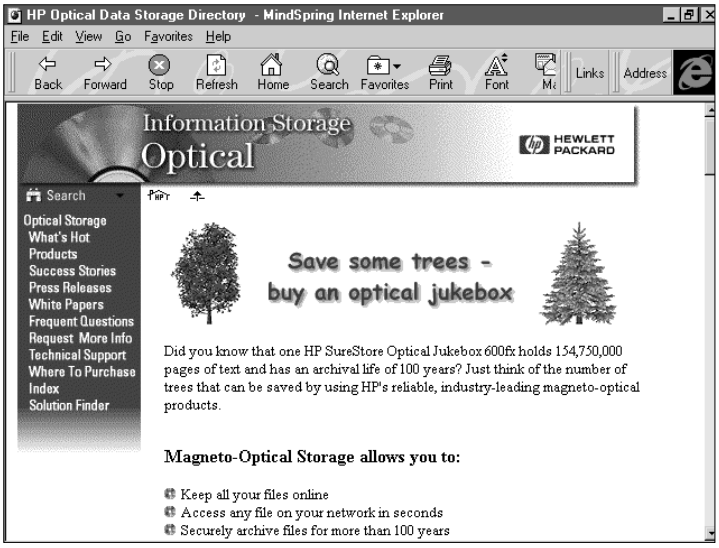


Figure 4.18 *Exploring optical storage at Hewlett Packard.*

Among the products listed are HP SureStore Optical Jukeboxes (of various sizes and storage capacities), HP SureStore Optical Multifunction Disk Drives, and HP Magneto-Optical Disks.

Another major vendor to visit when exploring storage is IBM. From their Web site's *Storage* page (**http://www.ibm.com/storage/**), you can access descriptive information about the IBM Seascape storage enterprise architecture.

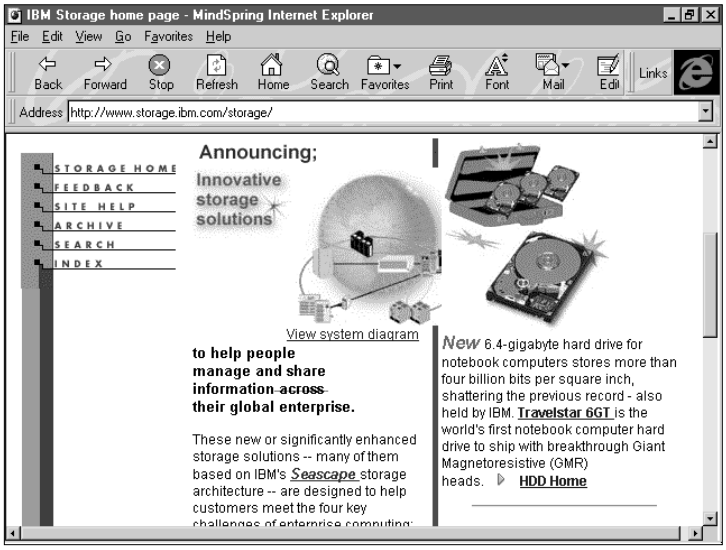


Figure 4.19
The IBM Storage Web page.

Described at the site as "a blueprint for comprehensive storage solutions optimized for a connected world," the Seascape architecture integrates various technologies (disk, tape, optical, etc.) to build reliable, versatile storage solutions. Visit the IBM *Storage* page to learn more about Seascape and other IBM storage options.

Much like IBM, Kodak (**http://www.kodak.com**) has taken a different approach to storage issues with their Digital Archive solution. As you remember, one of the primary benefits of DIP systems is the dramatic reduction (if not complete elimination) of the high cost of document storage. However, according to their white paper, "Digital Archive: A Strategic Positioning," Eastman Kodak contends that, "digital information is only as permanent as the hardware and software that gives it intelligibility. As technological evolution over even the past ten years has shown, there is little assurance that today's technology will be any more permanent than quadraphonic record albums which set the standard for audio capture and reproduction 20 years ago."

Certainly the digitization of business and personal records has raised the issue of how best to achieve long-term storage solutions; the possibility of storing precious data in a medium that could become obsolete in a few years is every manager's worst case scenario. The Eastman Kodak white paper puts it this way:

> *The relationship between digital archive and online storage reflects a marriage between speed and efficiency of access and attractive ongoing maintenance costs. It's generally conceded that retaining digital information online for extended periods of time is far from a cost-effective alternative. Maintaining digital documents actively online perpetuates a never-ending cycle of costly equipment expenditures needed to keep pace with expanding storage needs. Add to this labor and media costs related to ongoing migration and conversion to current compatible versions of computer hardware and software, and the online ante is upped significantly. To complicate the issue, the past has provided no model for the archiving of digital documents. Only now does digital archive, working with DDMS [digital document management systems], provide an option with the quick, transparent and cross-system retrieval capability of today's advanced document warehouses. Furthermore, digital archive is not vendor-specific; it can be used across a number of systems and platforms–as a stand-alone solution or as a component of a document warehouse system. Emerging digital archive technology is playing a key role in bridging the digital solutions gap. This is welcome news for document intensive organizations with long-term archival requirements such as insurance companies, banks, health care centers, hospitals, government agencies, and public utilities.*

Their proposed solution? The Document Archive Writer. This product works much like a laser line printer. Only, instead of printing to paper, Document Archive Writer commits digital

images to microfilm (at rates as high as 6.000 duplex pages per hour). The files produced are typically TIFF images, and according to Kodak, "This equation gives digital information a permanence that until now was accorded only by paper or traditional paper-based microfilmed applications." Although designed to use film to capture images for archiving, the system can also use film scanners to turn them back into digital forms. Visit the Kodak Web site to learn more about this and other Kodak storage options.

Selection Criteria

Clearly, your storage options are diverse and making selections can be overwhelming. Understanding a few of the options in advance, can help you ask sensible questions when approaching vendors and or consultants. In addition, consider the list of features below. Although listed as features offered by the Kodak Digital Archive approach, this list can act as a starting checklist for the basic functionality you should demand from any storage solution that you are considering for your DIP system.

- ***Longevity****—The ability to access and read digital documents in the future with little or no image degradation.*
- ***Inter-operability****—The ability to access documents (information) across a variety of differing digital document systems.*
- ***Cost****—Lower total life cycle cost viewed over the long term, including capital expenditures, storage media, operational expenses, maintenance contracts, and migration/conversion costs.*
- ***Integration****—The ability of the digital archive solution to work with other digital document system elements as an integral part of the records management process.*

- ***Technology Obsolescence Protection***—*The ability to exist through successive generations of new software and hardware.*
- ***Backup and Recovery***—*The ability to support business functions, if required, using archived documents.*

Developing Selection Criteria

Now that you have studied products offered by various DIP/EDM industry leaders, you are ready to begin to build some criteria for selection. The criteria that are most important to your organization will vary based on unique needs of your staff, customers, clients, and management. Use the worksheet that follows to help you identify questions that should be asked before and during your selection process. When researching various options, refer to this list to identify those products that seem to meet all of the needs you identified in your assessment. However, if no product seems to offer everything you need, don't despair. Most vendors work with organizations to customize and enhance DIP and EDM systems to meet the particular needs of their clients.

Choosing DIP Products

Functional Assessment:

In the space provided below, identify the tasks and functions that your DIP system must be capable of performing.

In the space provided below, identify additional tasks and functions that you would like for the system to perform, if cost allows.

User Assessment:

In the space below, describe the skill sets currently mastered by your organization's staff. (For example, are they *Windows* competent?)

Technology Assessment:

In the space below, identify any existing hardware components that you plan to use with your DIP system.

In the space provided below, identify your network platform and the operating system platform(s) that will be installed on workstation machines.

Cost Assessment:

In the space provided below, identify the amounts that have been budgeted for the purchase of new software and hardware.

Other Vendor Resources

When researching particular products, you have a wealth of tools on which to rely. Case studies, many of which are available online, are very valuable tools for understanding how organizations of various sizes and from diverse fields have made DIP software and hardware options work for them. In addition, computer and technology publications often review similar products from different vendors and their reviews can be a valuable source of information. Although you may not agree with the best product choices expressed by these publications, the criteria that they use for selection can help you identify product features that you may need to evaluate when considering various products for your own organization.

The growing popularity of the Internet as a research and commerce medium has also compelled many vendors and organizations to offer product information and white papers at their Web sites. These publicly accessible sites offer a wealth of information for the DIP project manager, including free downloads of analysis tools, demos, or trial versions of selected products. Chapter Eight of this manual is devoted to supplying addresses for a large listing of Web sites where case studies, product information, and other DIP resources can be found. Turn to the vendor listing in that chapter now if you would like to review other product vendors before proceeding.

Executive Summary

In previous chapters, you learned what constitutes an electronic document management system in general and a DIP system in particular. Additionally, you learned about the various components of Document Image Processing. In this chapter, by reviewing various products and vendors, you have seen that DIP hardware and software options are as numerous and varied as the needs of various businesses. Additionally, you have learned that, as a member of a DIP project team, sifting through the various hardware and software possibilities can be a daunting task. But whether you opt for a modestly priced scanner/software package, or a high-end system with customized components, manufacturers and vendors are in fierce competition for your business. This type of buyers' market allows you to get the most from your DIP budget.

Chapter Five
Web-Enhanced Document Processing

Perhaps you've noticed. The World Wide Web and its related technologies are having an enormous impact on current business processes. And DIP is no exception. In fact, according to Sandi Smith, CPA, CMA, CDP,

> *Because the whole point of the Web is for people to view documents, it seems pretty natural to assume that the Web will be a perfect extension to imaging and document-management systems. Many companies think so too, and are setting up Web-based document-management technology in their businesses. (p. 46, 1997 Top Ten Technologies and Their Impact on CPAs, AICPA, Inc., 1997.)*

In general, Web-enhanced document management systems perform the same functions as those described earlier in this text: data capture, data indexing, data storage, and data retrieval. The difference is that in Web systems, the TCP/IP technology that supports the cross-platform flexibility of the world-wide Internet and intranet networks is used to support the processing of the business-critical documents managed by the system. In this chapter, we will describe how Web technology works, the benefits of these types of systems for DIP and EDM managers, sample products, and selected security issues. Topics that will be covered in this chapter include the following:

- TCP/IP and other important Web technologies.
- Benefits of Web-enhanced document management.
- Samples of Web-enhanced document management systems.
- Web security issues.
- Guidelines for decision-making.

Understanding Web Technologies

Before looking at examples of Web-enhanced document systems, let's spend some time learning the basics of Web technologies. Although we will not attempt to tell you everything there is to know about Web technologies, this overview should provide accounting professionals who are working on DIP planning teams with enough information to assess whether or not Web-enhanced document processing might be the answer for their organizations. In addition, with these basic concepts out of the way, CPAs will be well prepared to discuss the benefits and drawbacks of Web document systems with vendors, contractors, and any organization's IT professionals.

Transmission Control Protocol/Internet Protocol

As with any network, Web-based networks rely on special protocols, or rules of communication, to allow computers to effectively share data. The protocols used for network communications on the Internet, on many LANs, and on intranets, are TCP/IP. For the sake of simplicity, protocols are usually considered as a group, sometimes referred to as a "suite" or "stack," that includes a cluster of related protocols. Actually, the TCP/IP suite includes Transmission Control Protocol (TCP), Internet Protocol (IP), Address Resolution Protocol (ARP), Internet

Control Message Protocol (ICMP), User Datagram Protocol (UDP), and others.

TCP/IP networks range in size and complexity from a few small computers on internal networks to the actual Internet "core" network, consisting of large mainframe computers handling millions of bytes of data each minute. The actual physical connections between the various networks that make up the Internet take many forms, from T3 links (high-speed communications pipes that act as "backbones" between major locations and carry 45 MBPS loads of traffic), to SLIP/PPP connections that carry Internet traffic over high-speed modems through dedicated ports supplied by service providers.

Every site connected to a TCP/IP network has a unique Internet address, just like destinations on typical postal routes have individual addresses that make it possible for mail carriers to deliver letters to particular destinations. To send a message to another computer on a TCP/IP network, your message must be addressed to the person or machine to which you wish to send it, just as letters sent through postal mail must be addressed to the intended recipient.

Like most forms of sophisticated communication, TCP/IP relies on a layered model. Any electronic message sent through a TCP/IP system must use these different layers to successfully reach its destination. For example, to retrieve a Web page, you use the protocols of the Web client (browser) program on your computer to create a request for the Web page (this request is in the form of the document's address) and send it on to the TCP layer. This TCP layer is responsible for breaking the message into units known as packets that are small enough to move through the network without monopolizing resources, assigning identifying information in the header of each packet, and formatting the

packets into a common "language." Then TCP moves the message on to the IP level, which physically delivers the message. IP adds its own set of header information that is needed to route the message appropriately. Once the message has arrived at the correct address, IP turns the packets over to TCP again. Then TCP reassembles the packets and makes a connection with the appropriate application program–the Web server program on the computer where the document is stored. In most TCP/IP transmissions, the header information is stripped out when a transmission has been successful. As a result, this level of protocol and decision making is invisible to the typical user.

Clients and Servers

Another key factor in the usefulness of TCP/IP systems is the distributed processing technique that makes it possible for two computers to share the work load for a task–called client/server applications. By using high-speed communication lines and special distributed-processing software, processing tasks can be shared by computers both technologically diverse and spread over wide geographic areas.

As early as 1986, Sybase, a database software developer, used the term client/server to describe a software product that linked databases and applications. As business computing evolved, more flexibility was required than was possible using mainframes and complex centralized computer systems. So, the computing infrastructure that was used in business workplaces began to take advantage of the increasingly powerful PCs that were also developing to distribute the workloads across business systems between software programs residing on different computers.

Client/server computing is an extension of this approach to programming, called "modular." Modular programming assumes that separating pieces of large software programs into smaller

parts, called "modules," makes development and maintenance simpler by spreading the workload necessary to complete sophisticated computer tasks over more than one computer. The client module of a distributed system is a program that sends a message to a server program requesting that the server perform a task or service (such as sending a document). Client programs manage the user-interface portion of an application–sending requests, validating data, and displaying results. Client software also manages the user's interaction with local resources (such as keyboards). The server module of a distributed process fulfills the request sent by clients by performing whatever task is requested (such as file retrieval).

For example, FTP (file transfer protocol) is the procedure used to transfer files from a remote computer to a local computer. Using an FTP client program, a user can access a remote site's FTP server program and request files to download. In return, the FTP server retrieves the information requested from the file system and sends it to the FTP client software. The FTP client software then reads the information and displays it in a useable form.

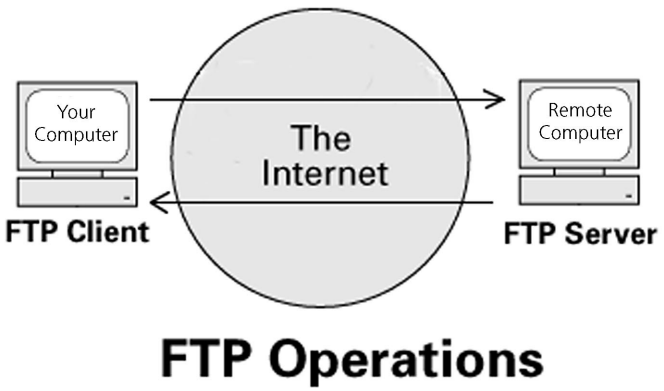


Figure 5.1 *A typical client/server application.*

Similarly, WWW client software, called a Web browser, allows a user to send a request to a WWW server in the form of a URL (Uniform Resource Locator) that is keyed into the browser. The server software at the WWW site responds by sending the requested document. The Web browser then reads the document's special coding (hypertext mark-up language) and displays it in a useable form on the user's screen.

The World Wide Web

When Tim Berners-Lee began working on his plan for the World Wide Web in the late 1980s, he was attempting, among other things, to solve a problem that continues to frustrate many businesses–using different workstations, procedures, and commands to access information on various systems. Berners-Lee reasoned that it must be possible to use a single interface to access all these different kinds of information. His proposal for dealing with this issue, "World Wide Web: Proposal for a Hypertext Project," paved the way for the creation of the World Wide Web and its supporting protocols.

The system that Berners-Lee envisioned was a client/server model. On the client side there was to be a Web browser, a common interface to display information that was neither platform-dependent nor proprietary–so that the already existing information servers could remain unchanged, for the most part. The key to the Berners-Lee model, then, was in the creation of a protocol running in the middle, Hypertext Transport Protocol, that offered a workable addressing scheme and a way for clients and servers to negotiate a document format.

Hypertext Transport Protocol

Designed to run as a layer on top of TCP/IP, Hypertext Transport Protocol (HTTP) is the core communication protocol between Web clients and servers. HTTP allows clients to send requests for

documents stored at specific server locations and to negotiate format presentation with the servers. By sending a list of the types of data the client understands and the data-transfer protocol to be used, the server can select a suitable format and use the correct transfer protocol. Locations and formats of objects on the World Wide Web are specified by Uniform Resource Locators (URLs). A URL specifies the server type (such as http://, gopher://, or ftp://), the location of the server (such as www.kentis.com), the name of a directory (such as */pub/*), and the file type (such as *filename.html* or *filename.txt*). For example, if you type the URL **http://www.kentis.com/design/design.html** in the URL box in your Web client software, you are requesting a service that meets these specifications:

http:// indicates that the document is stored on a WWW server.
www.kentis.com is an IP address that indicates the location of the server.
/design/ indicates a specific directory on the server.
design.html indicates a specific HTML (hypertext mark-up language) document stored on the server.

Hypertext Mark-Up Language (HTML)

The main mechanism for formatting and displaying Web documents is the system of formatting tags called HTML, which stands for Hypertext Mark-Up Language. HTML is derived from the Standard Generalized Mark-Up Language (SGML) that has been used for years as a means of open document interchange in industry and government. It is simply a series of tags that are included in a document to define qualities and formatting conventions, so that the document can be read across platforms and software types.

Although Web pages often seem sophisticated and complex, HTML is surprisingly easy to learn and implement. HTML tags are simple words or abbreviations enclosed in greater-than and less-than brackets (**< >**). These tags tell a Web browser how to treat the text within the tags–from defining appearance to creating links to other sites. Often, HTML tags are paired, surrounding the text that the tags will affect, but HTML tags never appear in the document that is displayed by a Web browser. For example, you can surround text with the paired HTML tags **<STRONG></STRONG>** to tell a Web browser to represent the text between the tags more strongly than the other text in the document. Most browsers will treat this tag sequence as a command to display the text between the tags in boldfaced lettering.

HTML documents can be created with simple text editors (like *Notepad* on a *Windows* PC), with any of many HTML editors currently available (many for free through Internet downloads), or through document conversion programs that convert documents created in various applications into HTML files.

Web Servers

Web servers are the central internal storage areas that manage and distribute information in response to client requests. In Berners-Lee's plan for the World Wide Web, Web servers have only one job–to respond to client requests. Servers do not need to know anything about other servers, or about the network on which they are running. Typically, at least one corporate Web server is used in an internal Web, but some organizations even choose to have multiple servers (at different geographic sites or for different departments, for example). Some organizations even use the same server for their intranet and public Internet documents, since most servers allow reasonable security by permitting the Webmaster to specify that certain IP addresses,

user names, or passwords are allowed or not allowed entry into designated areas.

Actually, Web servers have two parts: (1) hardware and (2) software. Almost any modern computer system equipped for networking can host a Web server. The hardware you select will depend on the traffic levels you anticipate, the ease of setup you require, and the technical background of your in-house staff. If you do not anticipate high traffic, you may find that dedicating a Mac or PC to be used as your server will meet the needs of your network. After you've selected hardware, you must also select server software to load on the host machine. Server software stores and distributes documents, and, as you might imagine, many different types of server software are available from various software vendors. If you choose to use an integrated suite of products, you may find server software included in your purchase, such as *PowerWeb Server,* offered with Westbrook Technologies' *Fortis PowerWeb* system.

Web Browsers

Web browsers do two things: (1) they act as viewers that display documents, and (2) they serve as client programs that send requests to a server on a network. As a viewer, a Web browser interprets the HTML tags in a file and displays formatted documents on the screen. The two most popular Web browsers available today are Microsoft's *Internet Explorer* and Netscape's *Navigator*. These two products remain in fierce competition for the browser market; the result is frequent advances and upgrades.

Web browsers can access and display a wide variety of documents and objects found on computers, including:

- HTML tagged text.
- Grayscale images.

- Black and white images.
- Full-color images.
- Video files.
- Sound files.
- Externally viewable files.
- Interactive forms.

Originally, only the formatted HTML text and compatible graphics were displayed within a browser's window. Other types of files automatically triggered the appropriate viewing software, which can read and playback other information formats like video, sound, and PostScript. However, enhanced with special software plug-ins that install on top of the basic browser software and expand the Web browser's functionality, browsers can now display other data types directly inside a browser window. One of the newest developments in this area is the Internet (or active) desktop in which all computing functions are displayed through the easy-to-use browser interface.

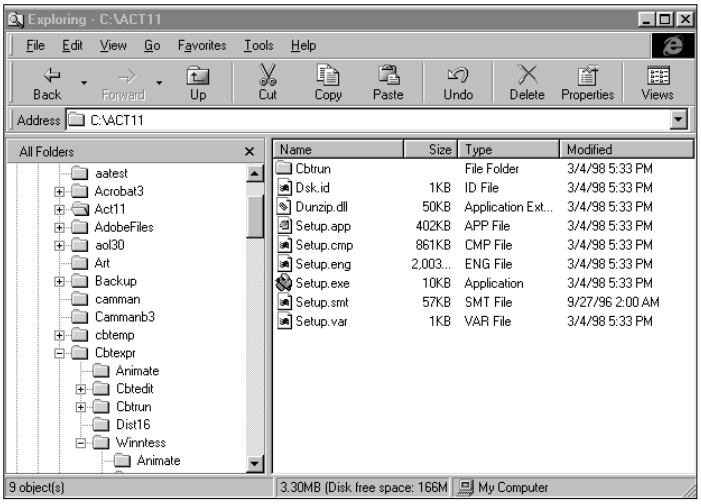


Figure 5.2 *Viewing a Windows Explorer directory listing through a Web browser interface.*

In addition, many Web-enhanced document management software suites use various types of programming to allow for automatic HTML conversions "on-the-fly." It is this recent enhancement in browser capabilities that is making document management through Web environments practical and desirable for many organizations.

Web Programming

As you explore various Web-enhanced imaging and document management tools, you will find that they use a range of programming strategies to extend the functionality of their baseline document processing systems to Web environments. The two most common types of programming that you will encounter are Java and ActiveX.

Java is a simplified version of the C++ programming language, designed to be securely and easily executed across networks. Using Java, programmers create small application programs, known as "applets," which run on the client side of a network. Applets are composed of code that makes it possible for clients and servers to exchange data, both dynamically and bi-directionally. In addition, by merging with World Wide Web technologies, Java makes it possible for programmers to create distributed, object-oriented applications (which treat programs as a collection of individual parts, or objects, that complete different tasks) that are not dependent on any particular desktop environment.

ActiveX is Microsoft's answer to the move towards greater interactivity on Web pages; it is a set of integration technologies that allows software components to work together in a networked environment. As you might imagine, this set of technologies is closely related to the *Windows* Object Linking and Embedding (OLE) features. Using ActiveX, Web developers can embed

ActiveX programs (known as "controls") right into their Web pages. When a client, using a browser that supports ActiveX controls, views the page, the controls are sent to the browser and the program automatically runs.

Although the particulars of these programs will probably matter very little to your decision-making, the added functionality that they bring to imaging and document management systems will rank high in your list of points to consider when you are considering Web-enhanced imaging.

Benefits of Web-Enhanced Systems

As mentioned earlier, Web-based systems offer all of the same benefits (and pitfalls) as other document management and DIP systems. As a result, you should review Chapter Two's discussion of benefits and pitfalls when evaluating any Web-enhanced document processing system. In addition, these systems offer a few advantages that are specific to the Web interface on which they are built. In this section, we will describe these benefits.

Compatibility with Intranets

Many organizations have seen the enormous benefits that can be gained from office networking. In addition, many organizations have recognized that the ease-of-use and cross-platform functionality of networking built on Internet technologies justifies developing internal networks on the model of the Internet. These internal webs are called intranets and they are quickly gaining ground as the network environment of choice in today's technology-wise businesses.

If your organization already has an intranet or has plans to implement an intranet in the future, Web-enhanced document processing may be the best option. For example, two years after initiating their corporate intranet, Dow Chemical Co. discovered that, although they had created a successful standard interface to a vast number of applications used by staff members, the number of documents, both scanned versions of documents and HTML documents, was vast and largely unmanaged. They turned to Documentum's Web-enabled software to bring document management to their intranet.

Remote Access for Staff

Whether or not your organization's internal network is Web based, you may be looking to the Internet for remote access options for your staff. If so, offering a Web interface for access to your organization's electronic archive may be the answer. Many vendors offer add-on products that can be used by remote staff members to access an organization's electronic archive. These products typically offer several levels of security to screen your company's valuable data from unauthorized users. For example, Plunkett & Cooney P.C., a 150-attorney law firm based in Detroit, is extending their document processing system, via the Web, to supply important documents to both home users and clients. Using *CyberDOCS*, an add-on to PC DOCS' (**http://www.pcdocs.com**) *DOCS Open* baseline system used by the firm, lawyers working at home can access files stored at the main office. (In addition, clients are given passwords that allow them to access files relating to their cases.)

Remote Access for Clients, Customers, and Business Partners

As organizations enter the World Wide Web in greater and greater numbers, the functionality of Web sites has been forced to grow. Until recently, WWW pages for many organizations were

little more than electronic billboards. But today, WWW pages are powerful interactive tools used to collect information from users, to market products and services, to support clients and customers, and to process transactions.

In organizations in which the Web has been recognized as the powerful business and communications tool that it is, selecting Web-enhanced document processing systems is the natural next step. When describing the benefits of using FileNET's *Panagon IDM Desktop* product, Richard Dillhoff, chief executive officer of InfoAccess (**http://www.infoaccess.com**) comments,

> *Instead of creating a new class of technology for managing the content of Web documents, IDM Desktop allows end-users to leverage their purchase of an integrated document management system by using it as a Web content manager. Panagon IDM Desktop customers can put complex compound document publications together easily; yet to a Web user, they look like one document complete with navigation aids, tables of references, and superior HTML usability.* ***(http://www.filenet.com/prods/panagon_info/Info_Access.html)***

For many organizations, this additional functionality can be a big factor in making your business case for a document processing system. The addition of Web capabilities that can extend to the Web objectives of your organization may make it easier to justify the expense of a document management system. As Murray Lorenzo, project leader at the Alberta Energy and Utilities Board expressed it, "It's that leap further, that extra bang for your buck," that can make the difference.

Ease of Use

The most obvious benefit of using Web-enhanced document management systems is that the user interface is easy to learn because it is built on tools that are familiar to most businesspersons. In some systems, the interface may simply be a common Web browser, such as *Navigator* or *Internet Explorer*. For example, the screen below shows a document displayed in an *Internet Explorer* window through FileNET's *Panagon IDM Desktop* product.

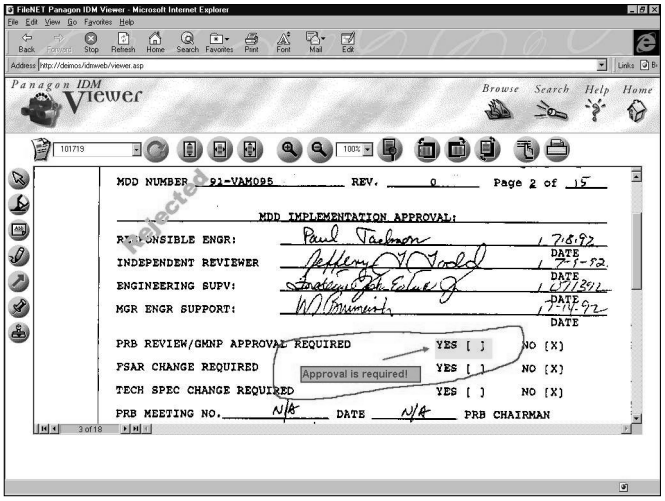


Figure 5.3 *Scanned document displayed in Internet Explorer through Panagon IDM Desktop.*

Other systems may blend screen elements from the document processing system with browser displays. Still other organizations may choose to have their systems built on a customized browser interface. In all cases, the simple point-and-click functionality of the Web environment can be leveraged to improve efficiency, reduce training costs, and increase staff confidence in the new system.

Examples of Web-Enhanced Systems

As you consider the possibility of a Web-enhanced imaging system, you will have many products from which to choose. In fact, in many cases, the Web components of document management systems are simply add-ons to self-standing EDM and imaging systems (as is the case with *CyberDOCS*). If your organization is already using some type of document management software, you may wish to contact your vendor first to determine whether or not Web utilities can be added to your existing system.

In the sections that follow, we will describe three popular Web-enhanced document management systems; there are many more! These descriptions should, however, give you a good basis for understanding the types of functionality that you can expect from these types of systems. As you read these descriptions, you may wish to make notes in the page margins about the features that seem most compelling and useful to your organization's current needs. Then, when you begin researching other product options, you can use these notes as your checklist for minimum functionality.

OpenText's Livelink Intranet

When I began my research for this section of the text, I solicited recommendations from EDM and DIP professionals in various discussion groups and online forums. *Livelink* was, by far, the Web-based document management program most frequently praised and recommended by these IS and IT professionals. If you visit the OpenText Web site (**http://www.opentext.com**) and view the *Livelink* demo, it is easy to see why. *Livelink* offers a clean, integrated, Web-based work environment that provides easy access to a searchable archive of various document types, as well

as workflow, project planning, and team management tools. Although the capture of documents happens outside the *Livelink* environment, the archiving and management of those documents, whether they are word-processed documents or digitized images, occurs in a Web interface. The user's experience is likely to begin on an intranet home page, such as the home page shown in the *Livelink* demonstration at the OpenText site.

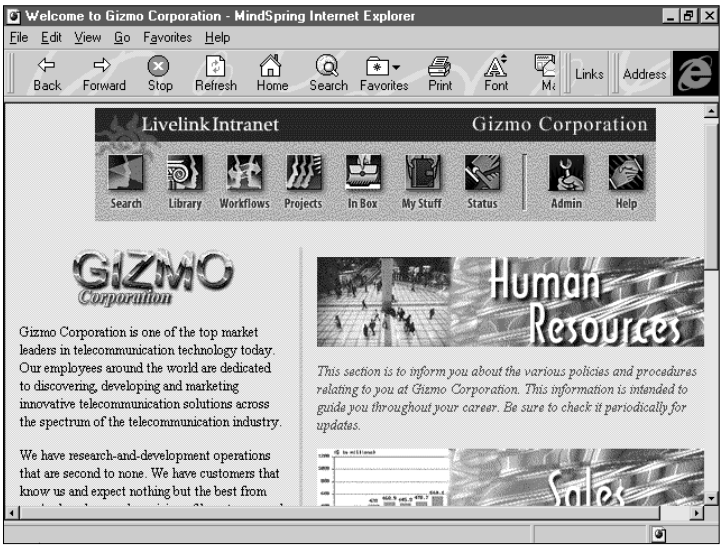


Figure 5.4 *The Gizmo Corporation home page.*

Documents can be easily uploaded to the *Livelink* library, and, from there, users can search for documents that are stored in the library, using a variety of search criteria and fields.

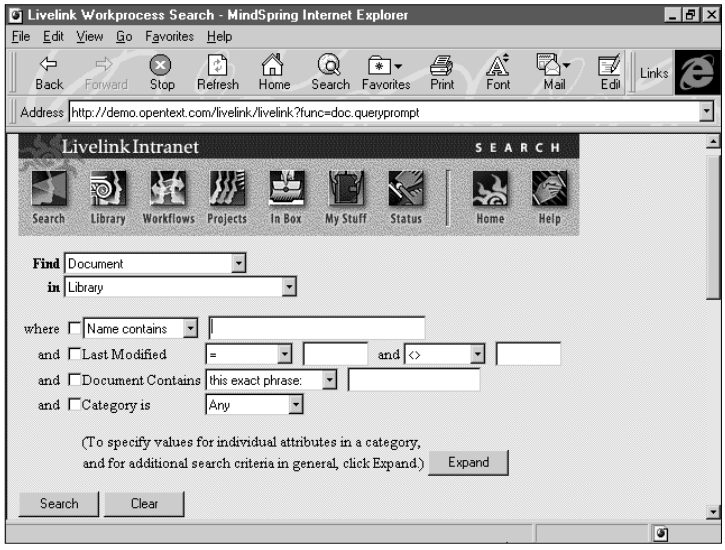


Figure 5.5 *Searching for documents in the Livelink library.*

When a document is located, it can be opened using either the applications on the user's computer or a Web browser interface. Most importantly, the choice is made automatically; that is, if a user accesses a document that is created in an application that is available on her machine, then *Livelink* automatically detects and launches that program. On the other hand, if she accesses a document that was created in an application that is not available on her computer, *Livelink* converts the document to an HTML-viewable file "on the fly" and displays it in the Web browser window.

The important point to note here is that no manual HTML conversion is needed. When a document has been entered into the system, whether it is a simple electronic document created in a word processor or a document that has been stored as a digitized image through scanning, the program displays it to the user with equal ease. The differences in document types becomes largely a matter of indifference. And documents can be made immediately accessible!

Library documents can also be "checked-out" for viewing and editing. Project leaders can check on the availability of documents for which they are responsible by accessing the system's various administration tools, also displayed in a Web browser interface. For example, selecting the **Checked-Out Items** option from the administration options, opens the screen below, from which project leaders can reserve documents and note which team members have accessed various project documents.

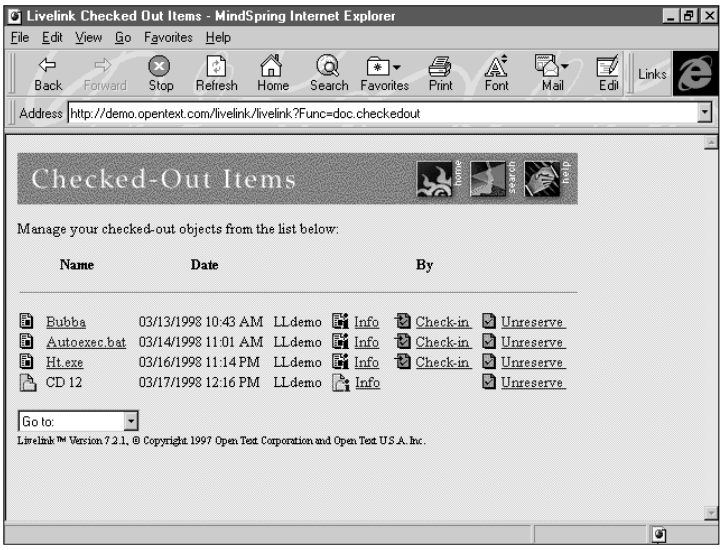


Figure 5.6 *Viewing the availability status of a library document.*

Westbrook Technologies' Fortis PowerWeb

Built on Westbrook Technologies' (**http://www.filemagic.com**) powerful *Fortis* product, *Fortis PowerWeb* provides all of the document processing functions of *Fortis* (document capture, document indexing, document searching and retrieval), through either a Web interface or a combination of *Fortis* and Web browser screens. As with other Web-based network software, the *PowerWeb* experience begins from a home page from where various more specific system areas can be accessed. After selecting a specific database, users must log on to gain access to

database documents that have been collected via scanning, electronic applications, fax, the Internet, etc.

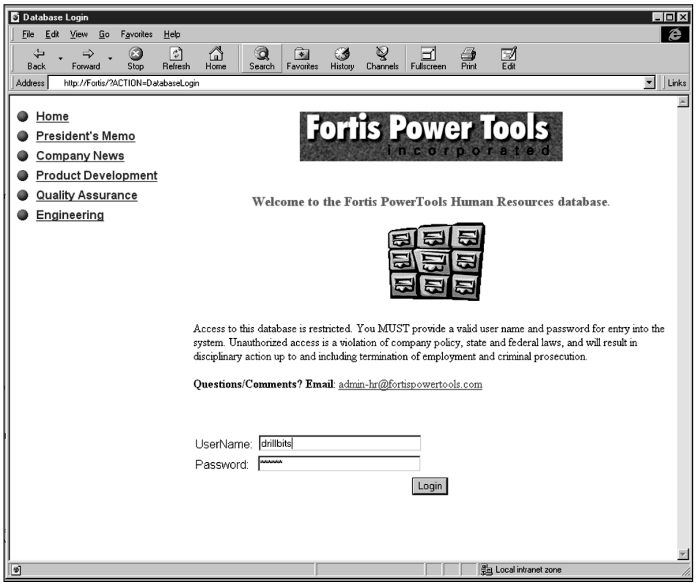


Figure 5.7 *Logging on to the PowerWeb document management system.*

After logging on, users can access documents by navigating through folders or launching a search using a Web-based search form.

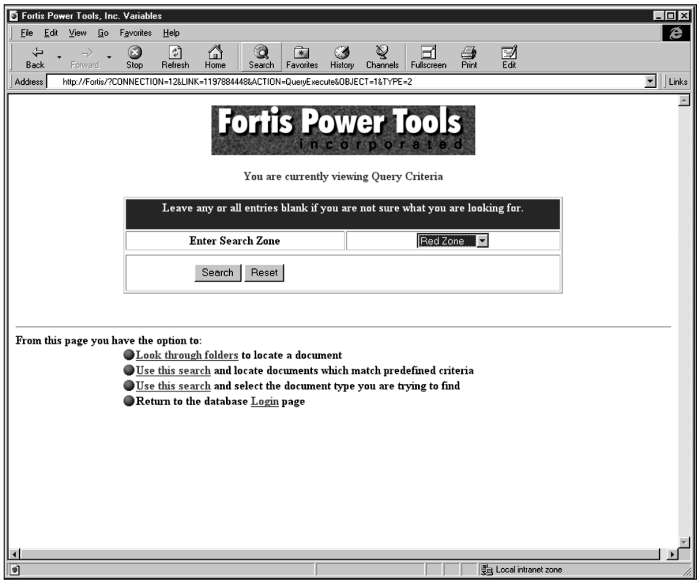


Figure 5.8 *Launching a query in PowerWeb.*

After a user enters criteria for the search, *PowerWeb* searches the selected database for matches and uses the powerful plug-in viewing tool to display the document within the same Web interface that the user has been working in from the start of the network session. The necessary conversion of the files is handled automatically by an active *Fortis* database. For example, the screen below shows the results of a search for a resume displayed in *PowerWeb*. The toolbar above the document allows users to control the document display and output easily from within the Web browser window.

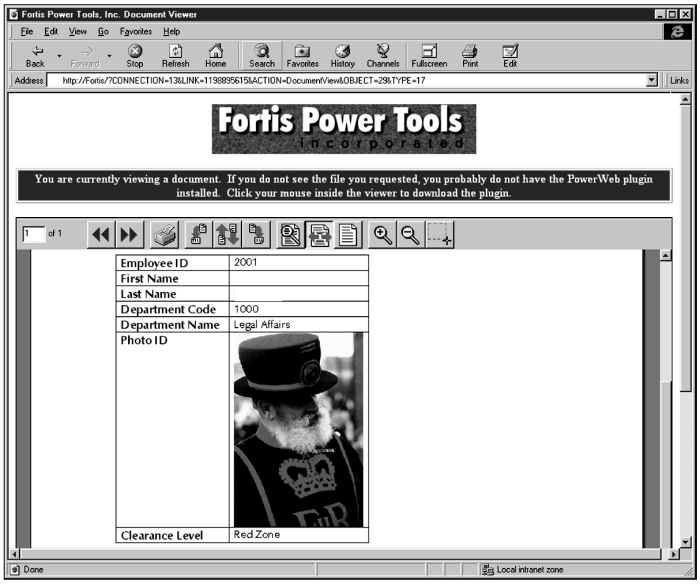


Figure 5.9 *Viewing a database document in PowerWeb.*

Although the screens shown above display *PowerWeb* at work with *Windows Explorer,* it also integrates with *Netscape Navigator*. The screen below shows a spreadsheet displayed in a *Navigator* window.

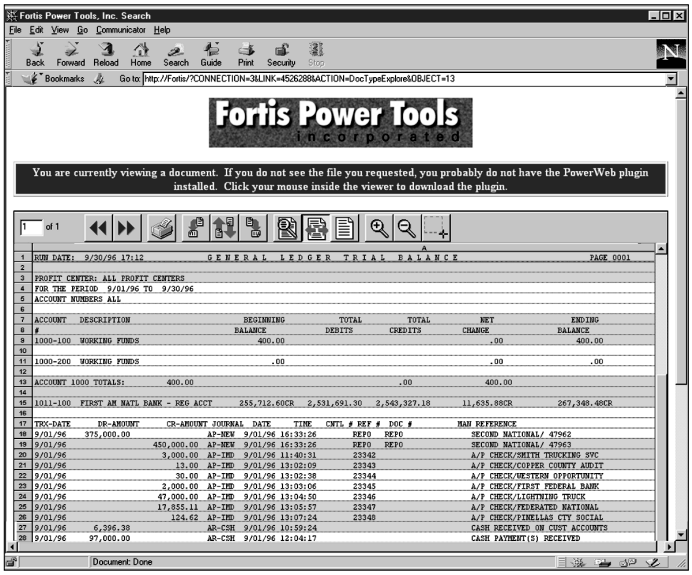


Figure 5.10 *Spreadsheet display in PowerWeb Netscape Navigator.*

Learn more about *PowerWeb* at the Westbrook Technologies Web site at **http://www.filemagic.com**.

Intranet Solution, Inc.'s Intra.doc

The *Intra.doc* document management system, is Intranet Solution's (**http://www.intranetsol.com**) answer to Web-based document processing. Like other Web-based document management systems, *Intra.doc* allows users to capture and store documents to a central archive or library. For example, the screen below shows the form that *Intra.doc* users would use to "Check-in" a *Word* document that is to be made available to users in a particular security group.

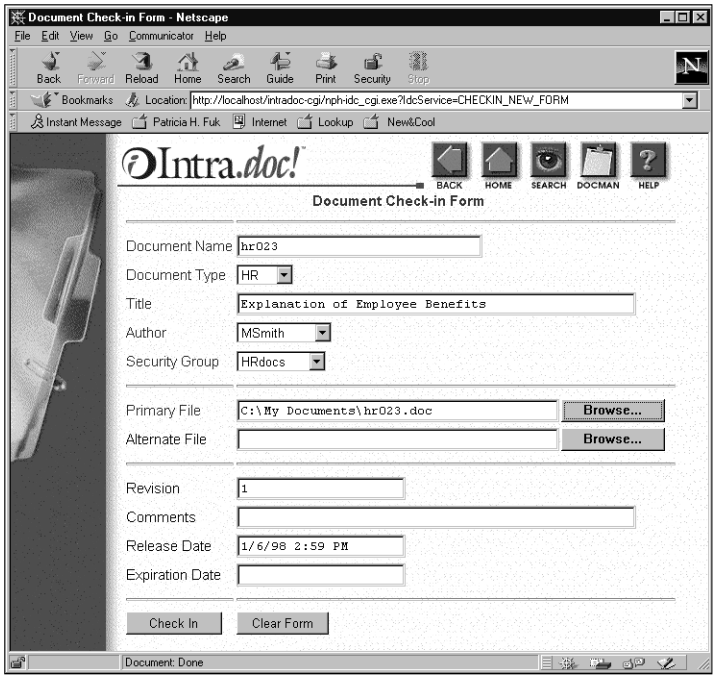


Figure 5.11 *Checking in a document in Intra.doc.*

Combined with the *Legacy* component, *Intra.doc* offers a different approach to Web-enhanced image processing. Through the *Legacy* software interface, users can scan, convert to a Web-ready format, and store documents in the *Intra.doc* electronic archive. *Legacy* organizes documents, electronic and paper, and converts them into the application-neutral PDF format described in Chapter Three. Users then launch a Web browser, with *Acrobat Reader* capabilities, to locate and view documents and images, regardless of their original file types.

Another feature offered by *Intra.doc* and other Web-enhanced document processing systems is the built-in workflow function. For example, the split screen below shows a user's electronic in-basket and a form in which a workflow is being initiated.

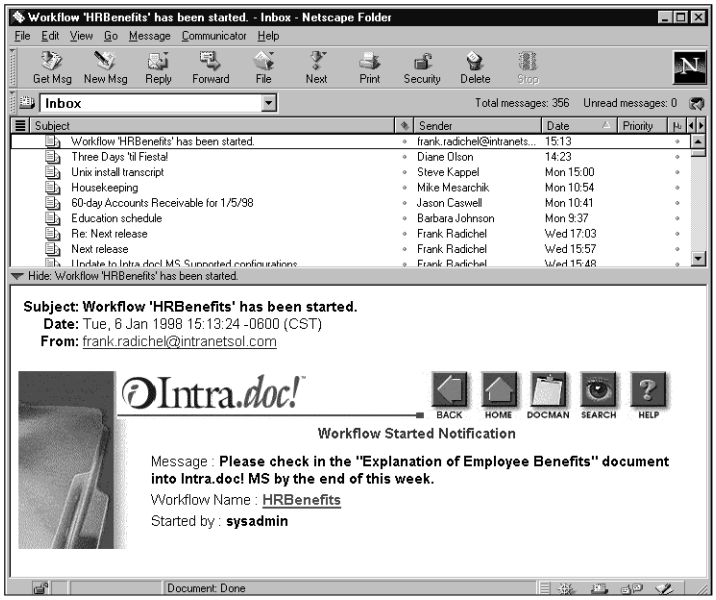


Figure 5.12 *Initiating workflow processes in Intra.doc.*

Obviously, the approaches used by these three vendors are quite different. Use these general descriptions to begin to assess the possibilities for Web-enhanced imaging and document management in your organization. (For up-to-date pricing information, visit the Web sites provided in the product descriptions.)

Web Security Issues

The convenience of connecting your DIP system to the Internet brings with it new security risks, since it opens your system to millions of potential contact points. In fact, many organizations have opted not to take advantage of the enormous benefits of connecting corporate networks to the Internet because of these risks. As with any new technology implementation, it is important to embark on your Web project armed with as much information

as possible. The fact is, most security risks can be minimized with a little planning.

In Chapter Seven of this text, we will describe various security and administration issues that are common to all DIP systems. But, in this section, we will focus our attention on the risks that are particular to Web-enhanced networks that are connected to the Internet.

As you know, the Internet is made up of a wide variety of hosts, from supercomputers to personal computers, including every imaginable type of hardware and software. And TCP/IP (Transport Control Protocol/Internet Protocol) is the "language" of the Internet. Anything that can learn to "speak TCP/IP" can communicate on the Internet. If you decide to connect your DIP system to the Internet, you will need to be concerned with TCP/IP security issues. In this section, we will review the risks and benefits of TCP/IP servers operating on the Internet from the standpoint of the TCP/IP protocol itself.

One of the most important features of TCP/IP isn't a technological one: the protocol is an "open" protocol, and anyone who wishes to implement it may do so freely. Engineers and scientists from all over the world participate in the IETF (Internet Engineering Task Force) working groups that design the protocols that make the Internet work. Their time is typically donated by their companies, and the result is work that benefits everyone.

IP

IP is a "network layer" protocol. This is the layer that allows the hosts to actually "talk" to each other. Talking to each other would include such things as carrying datagrams, mapping the Internet address (such as 10.2.3.4) to a physical network address (such as 08:00:69:0a:ca:8f), and routing, which takes care of making sure

that all of the devices that have Internet connectivity can find the way to each other. IP has a number of very important features that make it an extremely robust and flexible protocol. For our purposes, though, we're going to focus on the security of IP, or, more specifically, the lack thereof.

Attacks Against IP

A number of attacks against IP are possible. Typically, these attacks exploit the fact that IP does not provide a robust mechanism for authentication, which is proving that a packet came from where it claims it did. A packet simply claims to originate from a given address, and there isn't a way to be sure that the host that sent the packet is telling the truth. This isn't necessarily a weakness, per se, but it is an important point, because it means that the facility of host authentication must be provided at another point in the network configuration. Today, applications that require strong host authentication (such as cryptographic applications) do this at the application level.

IP Spoofing

This is an attack where one host claims to have the IP address of another. Since many systems (such as router access control lists) define which packets may and which packets may not pass based on the sender's IP address, this is a useful technique to an attacker: he can send packets to a host, perhaps causing it to take some sort of action. Additionally, some applications allow log in based on the IP address of the person making the request.

IP Session Hijacking

This is a relatively sophisticated attack, first described by Steve Bellovin. It is very dangerous, however, because there are now toolkits available in the underground community that allow otherwise unskilled crackers to perpetrate this attack. IP Session Hijacking is an attack whereby a user's session is taken over and

controlled by an unauthorized user. If the user was in the middle of creating e-mail, the attacker is looking at the e-mail, and then can execute any commands he wishes just as the attacked user would. The attacked user sees his session dropped, and may simply log in again, perhaps not even noticing that the attacker is still logged in and using the system.

TCP

TCP is a transport-layer protocol. It needs to operate on top of a network-layer protocol, and it was designed to operate over IP. (Just as IP was designed to carry, among other things, TCP packets.) Because TCP and IP were designed together, and wherever you find one you typically find the other, the entire suite of Internet protocols is known collectively as "TCP/IP." TCP itself has a number of important features that we'll cover briefly.

Probably the most important feature of TCP is guaranteed packet delivery. Host A sending packets to host B expects to get an acknowledgment back for each packet. If B does not send an acknowledgment within a specified amount of time, A will re-send the packet. Applications on host B will expect a data stream from a TCP session to be complete, and in order. As noted, if a packet is missing, it will be resent by A, and if packets arrive out of order, B will arrange them in proper order before passing the data to the requesting application.

This is well suited for a number of applications, such as a Telnet session. A user wants to be sure every keystroke is received by the remote host, and that it gets an acknowledgment for every packet sent back, even if this means occasional slight delays in responsiveness while a lost packet is re-sent or while out-of-order packets are rearranged. It is not suited well toward other applications, such as streaming audio or video, however. In these, it doesn't really matter if a packet is lost (a lost packet in a stream

of 100 won't be distinguishable), but it does matter if it arrives late (because of a host re-sending a packet presumed lost), since the data stream will be paused while the lost packet is being re-sent. Once the lost packet is received, it will be put in the proper slot in the data stream, and then passed up to the application.

Firewalls

Connecting an organization to the Internet provides a two-way flow of traffic. This is clearly undesirable in many organizations, as proprietary information is often displayed freely within a corporate network. In order to provide some level of separation between an organization's LAN (on which the DIP archive and all its valuable data resides) and the Internet, firewalls may be employed. A firewall is simply a group of components that collectively forms a barrier between two networks. A number of terms specific to firewalls and networking are going to be used throughout this section, so we'll review them first.

- **Bastion host**—A general-purpose computer used to control access between the internal (private) network and the Internet (or any other untrusted network). Typically, these are hosts running a version of the *Unix* operating system that has been customized in order to reduce its functionality to only what is necessary in order to support its functions. Many of the general-purpose features have been turned off, and, in many cases, completely removed, in order to improve the security of the machine.
- **Router**—A special purpose computer for connecting networks together. Routers also handle certain functions, such as routing, or managing the traffic on the networks they connect.

- **Access Control List (ACL)**—Many routers now have the ability to selectively perform their duties, based on a number of facts about a packet that comes to it. This includes things like origination address, destination address, destination service port, and so on. These can be employed to limit the sorts of packets that are allowed to come in and go out of a given network.
- **Demilitarized Zone (DMZ)**—The DMZ is a critical part of a firewall. It is a network that is neither part of the untrusted network nor part of the trusted network. But this is a network that connects the untrusted to the trusted. The importance of a DMZ is tremendous–someone who breaks into your network from the Internet should have to get through several layers in order to successfully do so. Those layers are provided by various components within the DMZ.
- **Proxy**—This is the process of having one host act on behalf of another. A host that has the ability to fetch documents from the Internet might be configured as a proxy server, and hosts on an internal intranet might be configured to be proxy clients. In this situation, when a host on the intranet wishes to fetch the **http://www.megasoft.com/** Web page, for example, the browser will make a connection to the proxy server and request the given URL. The proxy server will fetch the document and return the result to the client. In this way, all hosts on the intranet are able to access resources on the Internet without having the ability to directly connect to the Internet.

Types of Firewalls

There are three basic types of firewalls, and we'll describe each of them briefly below.

Application Gateways

The first firewalls were application gateways; they are sometimes known as proxy gateways. These are made up of bastion hosts that run special software to act as proxy servers. Clients behind the firewall must be proxitized (that is, must know how to use the proxy and be configured to do so) in order to use Internet services. Traditionally, these have been the most secure types of firewalls, because they don't allow anything to pass by default, but need to have the programs written and turned on in order to begin passing traffic. These are also typically the slowest, because more processes need to be started in order to have a request serviced.

Packet Filtering

Packet filtering is a technique whereby routers have ACLs (access control lists) turned on. By default, a router will pass all traffic sent through it and will do so without any sort of restrictions. Employing ACLs is a method for enforcing your security policy with regard to what sorts of access you allow the outside world to have to your internal network, and vice versa.

There is less overhead in packet filtering than with an application gateway. Due to the lower overhead and the fact that packet filtering is done with routers, which are specialized computers optimized for tasks related to networking, a packet filtering gateway is often much faster than its application layer cousins.

Because this type of firewall works at a lower level than an application gateway, supporting new applications either comes automatically or is a simple matter of allowing a specific packet type to pass through the gateway. There are problems with this method, though. Remember, TCP/IP has absolutely no means of guaranteeing that the source address is really what it claims to be. As a result, we must use layers of packet filters in order to

localize the traffic. We can't get all the way down to the actual host, but with two layers of packet filters, we can differentiate between a packet that came from the Internet and one that came from our internal network. We can identify which network the packet came from with certainty, but we can't get more specific than that.

Hybrid Systems

In an attempt to combine the security of the application layer gateways with the flexibility and speed of packet filtering, some vendors have created systems that use the principles of both. In some of these systems, new connections must be authenticated and approved at the application layer. Once this has been done, the remainder of the connection is passed down to the session layer, where packet filters monitor the connection to ensure that only packets that are part of an ongoing (already authenticated and approved) interaction are being passed.

Other possibilities include using both packet filtering and application layer proxies. The benefits here include providing a measure of protection for your machines that provide services to the Internet (such as a public Web server) as well as providing the security of an application layer gateway to the internal network. Additionally, using this method, an attacker, in order to get to services on the internal network, will have to break through the access router, the bastion host, and the choke router.

Making the Selection

Lots of options are available, and it makes sense to spend some time with an expert, either your in-house network administrator or an experienced consultant, who can take the time to understand the information security needs of your DIP system, and can design and build a firewall architecture that best

implements those needs. Other issues like services required, cost, and upgrade options might factor in to the final design.

The business of building firewalls is in the process of becoming a commodity market. Along with commodity markets come lots of folks who are looking for ways to make a profit without necessarily keeping their customers' best interests in mind. Additionally, vendors compete with each other to try and claim the greatest security, the greatest ease of administration, and the least visibility to end users. In order to try to quantify the potential security of firewalls, some organizations have taken to issuing firewall certifications. The certification of a firewall means nothing more than the fact that it can be configured in such a way that it can pass a series of tests. Similarly, claims about meeting or exceeding US Department of Defense "Orange Book" standards, C-2, B-1, and others, all simply mean that an organization was able to configure a machine to pass a series of tests. This doesn't mean that it was loaded with the vendor's software at the time, or that the machine was even usable.

Executive Summary

In this chapter, we have overviewed the technologies that support Web-enhanced DIP tools and the benefits and risks of implementing Web-enhanced DIP and document management systems. In addition, we've looked at some products that approach Web-enhancement of document management in diverse ways. Finally, we've reviewed a few of the specific security issues that accounting professionals acting as DIP project leaders and managers must consider when planning for Web-enhanced DIP. Use the worksheet on the following page to help you determine whether or not Web-enhanced DIP is the best solution for your organization.

Do You Need Web-Enhanced DIP?

Apply the questions below to your organization. The brief discussions that follow each question will help you consider the implications of your answers.

yes/no

Does your organization have an intranet? _____
If your answer is "yes," chances are you will want your document and image processing tools to use the same easy-to-use interface and cross-platform flexibility that led you to decide on Web-based networking in the first place.

Does your organization plan to have an intranet? _____
If you answer "yes" to this question, you will want your document and image processing tools to integrate easily with your intranet when you make the conversion. You may even use the Web-enhanced DIP software as an opportunity to introduce the staff to completing their daily tasks in a Web-based environment.

Does your organization have a publicly accessible Web site? _____
If your organization has a Web site, you may wish to consider using a Web-enhanced DIP system that will allow you to reduce the time and resources spent converting documents into HTML formats that can be viewed at your Web site. The same simple, automatic tools used to make various types of documents available for internal viewers can be used to create documents that can be made available at your Web site.

Is your Web site used to support business-critical processes? _____
If you support business-critical processes through your Web site (such as purchase orders and customer support), using a Web-enhanced DIP system can help to integrate those Web site processes with other business processing, allowing for improved service and parallel processing.

Are your staff members comfortable with Web software? _____
In most organizations today, staff members are familiar with the WWW. The simple point-and-click strategies and use of buttons and hyperlinks can be quickly and easily mastered to reduce the amount of training time that you will need to invest in staff members when your new Web-enhanced DIP system comes online.

Chapter Six
Planning Your DIP Project

Information technology projects, such as the implementation of a DIP system, can require considerable investments in time and money. Careful planning at the start of, as well as during, a project can help to ensure that your DIP project will accomplish its goals on schedule and within your planned budget. You may be tempted to get started without considering where the project will lead; but decisions made in haste may need to be reversed later, may prove too costly for your organization, or may lead to discouraging results that leave management and staff members doubting the value of your DIP project.

To avoid these disappointments, use a well-defined planning methodology. A methodology is a proven set of steps and tasks that system developers follow to build quality systems quickly, at low costs, and with minimal risk. For example, one popular methodology is the LifeCycle methodology that sets goals, defines target dates and results, establishes schedules, and estimates costs for the entire life cycle of a project. If your company has its own successful, pre-established methodology, or you have successfully used some other method to implement information technology projects in the past, then select the method that best suits your organization's needs. The exact methodology that you choose is unimportant as long as it is appropriate to the size, type, and scope of your project. Whichever method you choose, be prepared to adapt your methodology to the particular challenges of implementing your DIP project by studying this chapter. Key topics include the following:

- DIP teams.
- Needs analysis.
- Site and resource analysis.
- Costs/benefits analysis.
- Document analysis.
- System-use policy.
- The pilot project.
- Tests and adjustments.
- Full implementation.
- System support.

Before reading the rest of this chapter, you may wish to review the Chapter One case study of New York Central Mutual Fire Insurance Company. We will refer to the experiences of this company at various points in our discussion of planning; having the case study fresh in your mind will help you to place these planning issues in a real-world context that can easily be extended to your own organization.

The DIP Team

Perhaps the most important aspect of planning and implementing a DIP project is selecting staff members and/or outside consultants for the DIP project team. No single decision will have as far-reaching an impact on the success of your project as the commitment and expertise of the people who help you plan and implement it. As a result, it is important to select team members carefully and to allow team members to use their diverse talents to bring the DIP system to fruition.

As with various other types of DIP decision making that we have discussed in this text, the size of your DIP team and the particular

team members will depend a great deal on the size of your organization, your DIP goals, and the technology background of your internal staff. As we mentioned earlier in this text, you may choose to supplement your project team with outside experts or you may need to invest in specialized training for various in-house staff members who will be made responsible for the technical decisions (such as hardware and software testing, selection, and implementation).

It is also important to remember that the people who can best describe the flow of documents and the various ways that documents are used are the staff members who use those documents eight hours a day, five days a week. Whether or not those staff members know anything about DIP technologies, their input on the development of your project will be instrumental in building a system that truly supports the needs of the staff members who will use it.

When planning your project team members, do not overlook the need for a documentation specialist. When selecting a system, you will need to document the problems you are trying to solve, the options you considered, and the reasons you selected and rejected certain approaches. You will also need to document controls and policies that are put into effect for each stage of document processing. These policies may become part of a general system-use policy later. If filling this position with an in-house staff member, choose someone with strong writing and organizational skills. It is generally a good idea to ask one staff member to coordinate the documentation for various team members as they complete project steps in parallel.

Finally, in addition to the representatives of various business processes and systems, staff members who can manage the nitty-gritty technical aspects of the system, and a documentation

specialist, the team will need a manager who will oversee the project from start to finish. As other staff members complete various tasks, they may drop from the team and return to their other responsibilities. On the other hand, the project manager's task is never complete because once implemented, the system will require continuous maintenance and support. You may remember Steve Cembrinski's comment about his role as "vision keeper" for New York Central Mutual Fire Insurance Company's project:

> *If you move into a project like this, you have to have one person who can maintain a hands-on relationship through the entire process and who can pay close attention for a long period of time. When one person is dedicated to the project, it is less likely the ship will stray from course. We wanted a variety of individuals on our team, but we found it was a good idea to maintain constancy of leadership within the project team.*

Developing a Written Plan

When embarking on an IT/IS project, it is important to develop a written plan that outlines the steps that the project team will complete to bring the new system online. Although the plan may require adjustments as various phases and tests are completed, taking the time to outline the project goals and your strategies for achieving them can give direction and focus to what may seem an overwhelming and vague long-term objective. Your written plan is likely to include many components that are uniquely geared towards the needs of your organization. Even so, the basic planning items listed and described in the sections that follow should be addressed somewhere in your written plan.

The format of your plan may be based on a template used by your organization or a project-planning software product, such as *Microsoft Project*, displayed below:

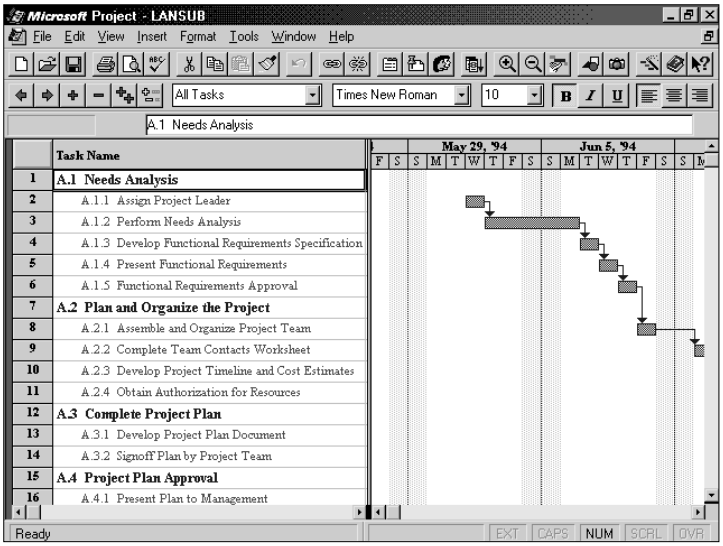


Figure 6.1
Using Microsoft Project to plan project goals and timelines.

Whatever approach you take, careful planning now will save you from headaches and delays later.

Needs Analysis

DIP systems fulfill two major organizational needs: (1) the need to archive valuable data and (2) the need to make that data easily accessible to staff members. Your organization may have one or both of these needs. The only reason to be considering the implementation of a DIP system for your organization is that you have identified some problem that DIP and its associated technologies can solve, or some need that a DIP system can meet. You may be reading this guide because you have already identified some problem in your organization that DIP might solve; or, perhaps, by learning about DIP and document management, you have come to identify weaknesses in your

organization that you once believed were irresolvable. Whatever process has brought you to this point, it is important to stop before continuing and carefully consider your organization's communications processes, which will be redesigned by the implementation of DIP.

To identify the needs of your organization, spend some time evaluating and tracking the lines of communication. In their assessment of the needs of New York Central Mutual Fire Insurance Company, Cembrinski's team made this step a vital component of their plan:

> *We spent a solid week in which I went through the department with the claims people and learned specifically each of the jobs out there. They taught me the process, and they looked closely at the details of the system they knew. We took various types of occurrences and tried to detail the steps involved in processing, asking questions like, 'What happens to that paper? Who else needs to look at it?'*

When completing these steps for your organization, you may wish to develop a flowchart that documents the movement of data and information in your organization. You might even circulate a questionnaire to solicit staff feedback on the most pressing communications issues they are facing. Be certain that at the end of this information gathering process you can answer the following questions:

1. How is information distributed in your organization?
2. What documents are used most often by staff members?
3. How do staff members retrieve business-critical information?
4. Is information freely available to everyone, or are special staff members responsible for controlling information in different departments?

5. Do staff members distribute information to all other staff members who may or may not need the information immediately or at some future date?
6. Are these information sources then used, filed, forgotten, or discarded?

If you already have a network, you should also examine the ways that information is accessed by the network users. Doing so will help you plan for network upgrades and training that may be necessary to encourage your staff to use the DIP system efficiently. You can begin your assessment of the network with the questions listed below:

1. What information is available?
2. How is the information accessed?
3. Is it easy to use?
4. Do staff members use electronic materials?
5. If not, why?

If your existing network includes an auditing or logging program, meet with your network administrator and assess what documents or functions are used most often. Also, ask the network administrator what kinds of problems are most prevalent in the current network and what support questions she is most often asked. Knowing these facts can help you determine strategies to reduce the maintenance and support time that will be necessary in your new system.

Armed with answers to these questions, collect and document some benchmark figures for your organization's communications processes. Identify the time required to complete key tasks using your current system and your project-improvement goals. Having these figures documented in your initial plan will help you to

accurately assess the successes and weaknesses of your DIP system when it comes online.

Site and Resource Analysis

You may already have a network or document management system in place in your organization. Or, perhaps, your current office situation is non-networked and you will be starting from scratch. At the very least, you are likely to have some hardware components that will be used in your DIP system. In all of the situations described above, it is important to assess and diagram your organization's current computer environment. If you have an existing network, draw a diagram of the network as it exists. Your diagram should include all of the following:

- Computers.
- Hubs.
- Cables.
- Printers.
- Power outlets.
- Power backup units.
- Telephone lines and jacks.

If your organization does not currently have a network, examine the current positioning of all computers, printers, power outlets, phone jacks, etc. You may also need to research any building codes that might complicate the placement of cables or other components of your network. Essentially, you are creating an inventory of hardware that may be "recycled" in your DIP system. And you are planning for the positioning of new hardware that you may need to purchase. For example, many organizations place input devices, such as scanners, in secure rooms that can be accessed only by those personnel who have been trained to prepare documents (both common and confidential) for archiving. Likewise, storage hardware is often placed in secure

rooms, separate from other computers, and accessible only to the system administrator. On the other hand, shared peripherals, such as printers, must be placed in locations where they can be accessed by many staff members. When this initial site assessment and inventory is complete, you can then use the diagram to sketch in the locations of the new system components. For example, the diagram below illustrates the current computing environment and connection upgrade plan for a small accounting firm planning to implement DIP.

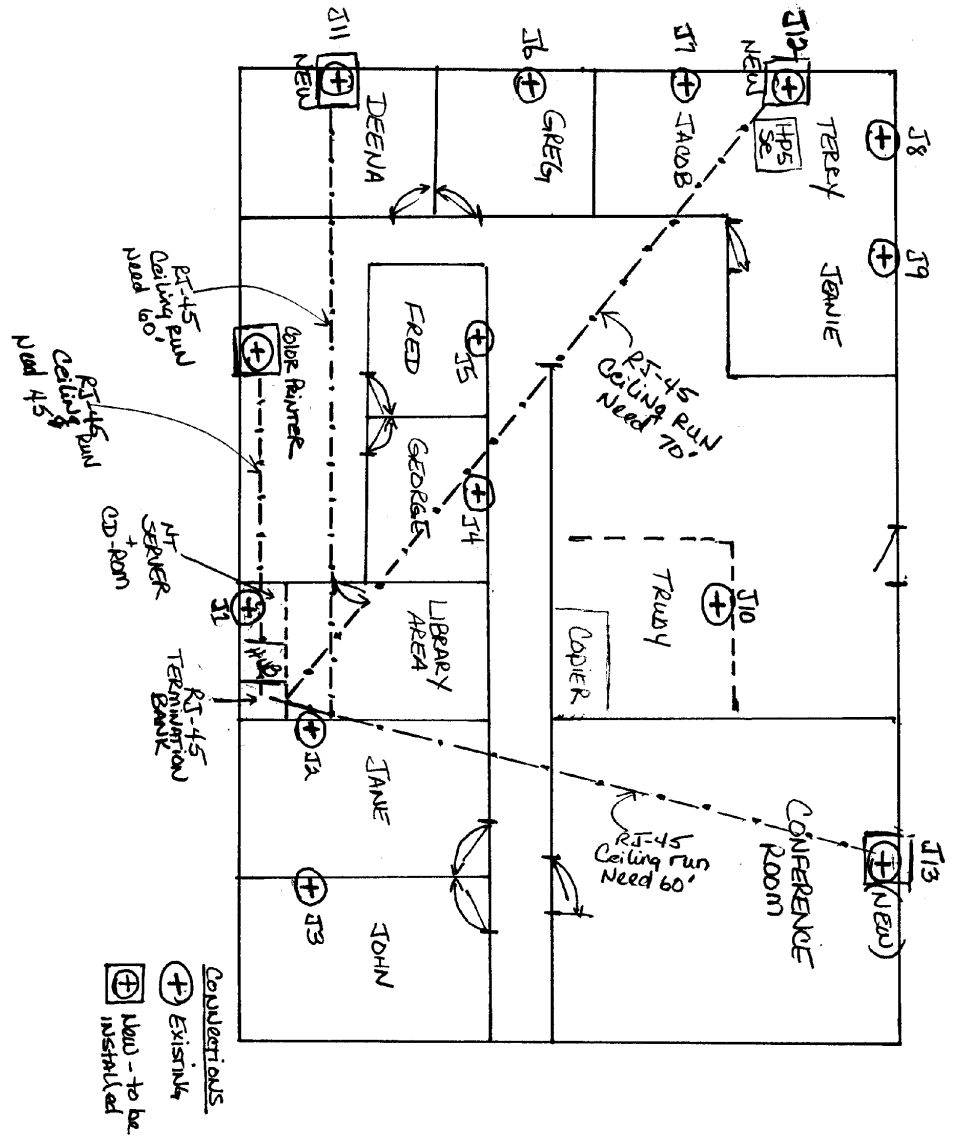


Figure 6.2
Sample site diagram.

In addition to assessing the computing environment in terms of hardware, you also need to assess the computing environment in terms of the people who use it. Specifically, you should plan to create a detailed listing of various staff members and the data to which they need to have access. Although the easy availability of information is at the heart of DIP and EDM systems, not all staff members will need access to all available documents. In fact, some documents will need to be restricted to certain high-level managers, such as payroll documents or confidential contracts. When you have determined the access and restriction needs of your organization, you can select software and hardware that will allow you to implement these security features quickly and easily. For example, many DIP and EDM products allow for documents to be archived in a way that allows access only to users who are members of certain pre-established user groups. Be certain to carefully consider and document the access and restriction needs of your DIP system before selecting any hardware or software solutions. (We will address access control and other security and administrative issues in Chapter Seven.)

Cost/Benefits Analysis

After you have determined the communications and resource needs that your organization is facing, consider the various solutions that can help you improve the flow of information and complete business-critical activities. Perform in-depth feasibility studies and cost/benefits analyses for several likely systems. Like Cembrinski, you may even choose to attend (or send team members to attend) various seminars hosted by product vendors.

Cost/benefits analyses are determined by comparing various alternative methods for achieving your organization's desired results (documented in the Needs Analysis section of your plan). Specifically, you must compare the costs of DIP functionality with similar functionality offered through other methods. In addition,

you must compare the DIP functionality offered by one DIP system or product suite to the functionality offered by other DIP systems and suites. You may also be considering the benefits and costs of implementing Web-enhanced DIP in this phase of your planning. At the very least, compare the top three contenders for the types of tasks you want your staff to be able to complete. This portion of your plan should consider each alternative's overall performance and factors such as:

- Compatibility with existing hardware and software.
- Adaptability to changing business needs.
- Impact on staff and organization.
- Reliability of the system.
- Ease of use.
- Costs.
- Service contracts from vendors.
- Possible upgrades.

For each alternative, prepare detailed cost estimates that include tangible factors as well as other intangible factors. The purpose of this type of cost/benefits analysis is to compare the technical, personnel, functional, and cost aspects of the various methods available for achieving your organizational goals. The resulting section of your project planning document will describe how the alternatives deliver (or fail to deliver) the results you expect. You may use the "Cost Estimate Worksheet" on the following page to gather useful data for developing this section of your plan.

Cost Estimate Worksheet

Item	Initial Cost	Continuing Costs
Network Backbone		
Installation/Licensing		
Upgrades		
Platform Upgrades		
Hard Drives		
Operating System		
Memory		
Hardware (capture, storage, output)		
Evaluation		
Installation		
Training		
Capture Software		
Evaluation		
Installation/Licensing		
Training		
Indexing Software		
Evaluation		
Installation/Licensing		
Training		
Storage Software		
Evaluation		
Installation/Licensing		
Training		
Retrieval Software		
Evaluation		
Installation/Licensing		
Training		
Maintenance		
System Upkeep		
Administration and Support		
Help Desk		
Continuing Training		

Document Analysis

Since the primary goal of your DIP system is improved document management, it is important to stop and consider the types of documents that your organization uses. Part of this step will be evaluating the levels of access that should be allowed or disallowed to the various types of documents that will be captured and stored by your new system. You may approach this step in any of several ways. For example, you may categorize documents in terms of the business functions and departments to which they are related (accounts receivable, accounts payable, human resources, etc.). Or you may categorize documents in terms of their access requirements (public access, low-level access, high-level access, etc.). You may also choose to use some blending of these two strategies, categorizing documents by business function and then by access requirements, for example.

Taking a close look at your documents in this early stage will help you to make important decisions about the security needs of your system and the personnel who will be assigned the task of organizing documents for capture and indexing. Obviously, only personnel with high-level security permissions can organize, scan, and index high-level security documents.

As you assess the documents that are used in your organization, you must also consider them in terms of their importance to daily business functions. As with the implementation of any business system, the transition to a DIP environment will take some time. Obviously, you will need to plan this aspect of implementation very carefully to avoid unnecessary service delays that will frustrate clients and customers. In part, this may require creative scheduling of input steps–for example, scanning a certain cluster of business-critical documents on a Saturday when staff members and customers are unlikely to be affected by their temporary unavailability.

System-Use Policy

There are a number of reasons for developing a system-use policy for your organization, reasons that are even more apparent if your system will also be open to remote access. Perhaps the most important consideration is productivity. The guidelines that you choose to provide will depend on your organization's objectives and staff policies, but it is a good idea to discuss these issues with management in advance and then develop guidelines to be distributed to staff members before the system is fully implemented. Using this method may prevent problems in advance, rather than leaving you in the position to clean up messes that are created later. Your policy should spell out the rights of individuals and draw clear lines between organizational and personal uses of the system's hardware, software, and information resources.

If you will be implementing a Web-enhanced system, you may also need to include Web style guidelines in your policy document. This will be especially important if individual staff members will be responsible for importing documents to the archive of electronic documents. These style guidelines must consider system issues. For example, defining guidelines for the number and size of graphics will allow you to control download times (and thus productivity) of the system. When making decisions about your guidelines, consult with your Webmaster and management.

The Pilot Project

After having progressed this far in the process, it will be very tempting to leap ahead with your DIP implementation–tempting but unwise. In fact, you may want to begin to develop your DIP system by focusing on a small manageable pilot project first.

Then you can build on that pilot project as the system tests are completed and adjustments made.

When you evaluated the communications processes in your organization, you probably noticed many likely areas for a DIP test. An obvious choice might be a document-heavy department, such as Sales or Human Resources. Or you might choose to implement the DIP pilot project in one office of a multi-office organization before implementing it in other offices. Cembrinski's team worked by territories, bringing one territory online, assessing feedback, and making adjustments before bringing on another territory.

Testing and Adjustments

As the pilot project is implemented, invite staff members to use the system and provide feedback. Be certain to follow up with staff members who are in charge of the information to find out how this new information distribution method affects their workdays. Are they spending less time answering staff, customer, and client questions? If so, how does this time saved compare to the time it takes to scan and archive the information? Use the feedback from this test phase to finalize the technical design, revise the performance model, and prepare system documentation. If the pilot project indicates problems you had not anticipated, then revise the project plan to offer solutions. Reassess the cost/benefits analysis you completed to make any necessary adjustments in your conclusions based on the real-life test run of your DIP project. Finally, meet again with your DIP team members to refine your development plan and then get ready for the challenge of full-scale implementation of your DIP system.

Full Implementation

Full implementation of your DIP system is not as imposing as it sounds, if you have followed the guidelines above to prepare for the project. In fact, this is the easy part, especially after you have implemented a pilot project on which you can build your expanded document management system. Moving users from your old system and onto the new system is a process known as "cutover." In many organizations, when a new system is installed, the old system continues to run during the initial test phase. This allows staff members to continue to complete their tasks without interruptions from the inevitable "tweaking" that will be necessary as you get the new system up and running. When you are ready for full implementation, you can add new users to the system in small groups. This will allow you to assess the impact of the growing workload on the various components of the system.

Depending on the size of your organization, this step may take a long or short time to complete. But even in small organizations, the cutover must be carefully planned to coincide with training and to cause the least amount of disruption to critical business processes. Obviously, this phase of the plan should include an outline of training strategies that will be used to bring your staff onto the system quickly. Many organizations plan for training to take place in controlled settings during the pilot phase of the network. This type of plan allows for staff members, newly trained, to make the cutover with very little disruption. If your office was not computerized before the implementation of your DIP system, you may find that you need to address two levels of training: basic computer training and specific training on your DIP system. Consult with your human resources staff to plan your training strategies and timelines well in advance.

System Support

After the DIP system is fully deployed, it is time to sit back and enjoy the fruits of your labor, right? Not necessarily. In most cases, full deployment must be quickly followed by the refinement of system support strategies. Your written plan should document the methods that you plan to use to (1) monitor users and support their needs and (2) monitor the system's performance and make necessary adjustments. Continuous, careful monitoring will also help you to identify and prevent potential problems, pinpoint where maintenance costs can be minimized, and determine when modifications are needed.

User Assessment

To successfully monitor your DIP system, you must establish and maintain a method of feedback for the system's users. User feedback is the most crucial information that you have, because, no matter how technologically smooth it may be, the DIP system is a failure if it is not actively and consistently used by the organization's staff. In addition to user satisfaction, include plans for assessing user performance. Your written plan will be most effective if you include benchmark figures for productivity gathered before the new system is installed. Describe the strategies that you plan to use to gather the follow-up statistics and the results that qualify as "success." Be as specific as possible; for example, if you are anticipating that the DIP system will improve customer support response times, gather benchmark figures and project an improvement goal. As you assess the working DIP system, these figures will help you draw realistic conclusions about the system's impact on your organization.

System Performance

Plan to use the feedback from end users, and logging and auditing software, to work with your system's administrator to eliminate system deficiencies. Again, gather benchmark figures

that you can use to measure the actual results against expected benefits and plan to respond continuously to system needs.

Executive Summary

In this chapter, we have overviewed the selection of DIP team members and various planning aspects that accounting professionals and financial managers working on DIP project teams should consider in the project planning document. Although your project plan may include other items that are unique to your organization, the points covered in this chapter must be addressed in some way to effectively plan for a smooth implementation of DIP systems in both small and large organizations. When you have reviewed the materials in this chapter, you may wish to use the worksheet on the following page to compile notes for your own DIP planning document.

Project Planning Worksheet

Plan Component	Responsible Team Members	Goal Dates and Notes
Needs Analysis		
Site and Resource Analysis		
Costs/Benefits Analysis		
Document Analysis		
System-Use Policy		
Pilot Project		
Tests and Adjustments		
Full Implementation		
System Support		

Chapter Seven
Security and Administration for DIP Systems

Clearly, DIP systems can result in rewarding benefits for organizations that implement them with planning and care. But, as with other IT systems, DIP systems also bring new risks that result, in part, from an organization's greater reliance on computers and automated systems. The risks that you will need to plan for fall into three broad categories: (1) physical security, (2) document security, and (3) end-user security. As we describe various examples, you will see that, in many cases, these risk areas overlap. Even so, managers who begin system risk assessments by categorizing risks in these three groups will easily identify key security issues for their own organizations. Major topics that we will address in this chapter include the following:

- Protecting physical components.
- Backup and recovery strategies.
- Preventing virus damage.
- Document accuracy.
- Document version control.
- Document authenticity.
- Passwords.
- User rights and permissions.

Physical Security

Many businesspersons think that information security is exclusively concerned with complicated computer code and system administration schemes. But, in fact, sophisticated software-based approaches cover only one aspect of securing information. Planning for the physical security of your organization's DIP system can be just as important. For instance, what good does a firewall do if someone can walk into your offices and walk out with the optical disks on which your critical data is stored?

Common sense is the best guide in office security. By placing storage hardware only in lockable rooms, you can gain a great measure of security. But, as you discovered when completing a site analysis for your system, not all DIP system components can be stored in locked rooms. For hardware that must be located in common areas, such as reception rooms or shared cubicle installations, office supply companies sell a variety of cables and plates that can be used to deter theft. Computer Security Products, Inc. (**http://www.ComputerSecurity.com/**), a company that provides such equipment, reports that the theft of computers now accounts for over $10 billion per year in losses.

As you plan for the physical security of your organization's DIP equipment, remember that computing equipment is a valuable asset and the data stored on this equipment is also an asset. Plan to protect your hardware, software, and data as rigorously as you would any other valuable company resource and begin with a simple assessment of the physical security of individual components.

Data Backup

Another aspect of planning for the physical security of your DIP system is to prepare for the worst-case scenario–loss of data through theft, sabotage, or system failure. Security experts who plan for the data security of large computer systems use complex terms, such as "volume dumps," "incremental dumps," "synchronization," and "geometric device differences." While you probably do not need to understand the nuances of these terms, you do need to have a plan for effectively recovering from lost or destroyed data.

Depending on the size of your organization, you may wish to consult with an IS professional who specializes in disaster recovery strategies. These professionals are trained to plan for the possible restoration of the contents of large data security devices, each of which can hold many gigabytes of information–often the data of organizations such as banks, healthcare facilities, and universities. In contrast, most data-loss incidents associated with small organizations are, relatively speaking, minor problems. When the hard drive of an employee's desktop computer workstation finally gives out, you can usually have a new one installed by the next office day. If a computer gets dropped during an office move, your insurance policy provides for its replacement. If a worker's notebook computer is stolen during a business trip, you can arrange for shared time on an office machine while a new one is purchased.

But what about the data lost during these incidents? If the optical disk jukebox on which your organization's electronic archive is stored is compromised, your DIP system could lose a good deal of important files. If you don't have backup copies of these critical files, what should be a minor inconvenience can become a major productivity setback, or worse. Organizations that implement DIP systems without providing for a regular,

procedure-driven backup system are flirting with disaster. In contrast, those organizations that design and implement rigorous backup systems are able to weather the small and large storms associated with being so dependent on those "boxes" on our desks.

When you think about data backup procedures for you organization's DIP system, you have two categories of decisions to make. The first is technology-dependent: what hardware and software will you use to make duplicate copies of the contents of your electronic archive? The second category of decisions is situation-dependent: what procedural system makes the most sense for your organization? Because data backup technology runs the gamut of prices, it might make sense to approach the technology category first by deciding how much your organization is willing to spend for reliable backups, and how much training you are willing to provide to employees. You should also explore any backup utilities that may be supplied with your system software. Then you can integrate all these factors into the data backup procedures.

The simplest and least expensive way to perform a backup of the contents of an individual workstation PC's hard drive is to use the backup function of the computer's operating system. Obviously, this procedure is not likely to be adequate for an integrated document management system in which files (frequently quite large) are stored in a central archive system, from which many users access important data. Next on the cost/complexity continuum is a group of hardware devices for backing up data. Still the most-used of such equipment, streaming tape drives have been very popular since the late 1980s and early 1990s. Designed as either internal devices (requiring a card slot) or external devices using a serial or parallel port, tape drives provide convenient storage of large amounts of data on high-volume

tapes formatted especially for the purpose. Tecmar Technologies (**http://www.tecmar.com**) is a large producer of tape backup hardware in several formats.

Figure 7.1 *Tecmar Technologies Web site.*

If your system is connected to the Internet, you might consider the services offered by some innovative Internet services companies. For instance, Sound Data Incorporated (**http://www.sounddata.com**) offers a service called *Intervault*, which combines client software for each workstation PC, software for the office's central server, transmission of the data via the Internet, and vaulted storage of the data for one fee. The data on each PC is backed up, gathered by the network server, encrypted, sent to the Sound Data server, and stored by the company for a specified interval.

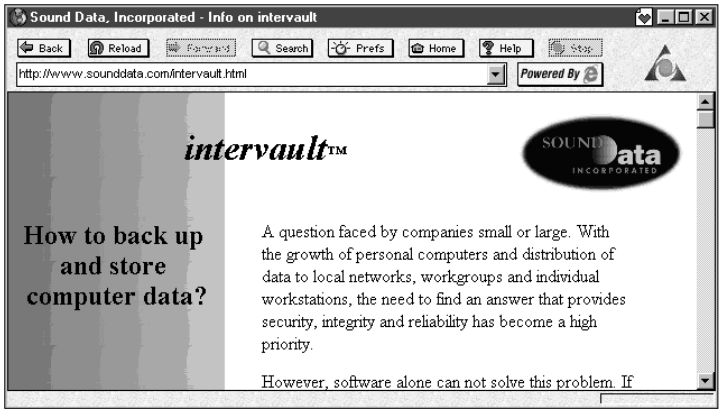


Figure 7.2 *Sound Data, Incorporated's Web site.*

Finally, do not forget the RAID (Redundant Array of Inexpensive Disks) technologies described briefly in Chapter Three. Using these systems of redundant disks and special error-correction software can help organizations recover data that is damaged due to system failures, disk degradation, viruses, or other types of damage.

Whichever software and hardware you choose to use, you will need to establish and follow a backup plan. Such a plan identifies the data that must be backed up, who is responsible for which tasks, and a schedule for completing backup procedures. For instance, you might choose to assign all backup responsibilities to the system administrator, who will make regular backups of stored files, following an iteration scheme that makes sense for your organization. An iteration scheme, incorporating off-site storage of at least one backup set at all times, is described below:

1. Each week all data files are backed up to streaming tape, which is sent to the off-site storage facility for safekeeping.
2. A second backup, from the previous week, is picked up at the off-site storage facility and transferred back to the office site.

3. A third set of tape cartridges, which were returned to the office site the previous week, are cleared by the system administrator and used for this week's backup.

As you see, this scheme allows for one complete tape backup of all files to be located off-site, one to be in process, and one to be in transit.

There are numerous online resources where you can learn more about data backup hardware and software options. For example, the Data Recall, LLC (**http://www.datarecall.com**) site allows you to view a multimedia demonstration of a service that is also usable with a stand-alone PC. By dialing directly to Data Recall on a modem with communication software, the PC can perform an automatic, encrypted backup of itself during off business hours. Another site that provides relevant information on data backup and storage is CyberStorage Systems, Inc. (**http://www.cyberstorage.com**). An industry leader in RAID technology, CyberStorage Systems, Inc.'s, site offers an overview of RAID technology, as well as current trends in RAID products and services

Figure 7.3 *Visit CyberStorage Systems' Web site to research RAID storage options.*

Computer Viruses

The very mention of the word "virus" causes managers, system administrators, and personal computer users to shudder. Almost everyone has a story to tell about a co-worker or friend who lost that irreplaceable report or crucial database due to a damaging virus. A recent National Computer Security Association survey reports that about 98% of the corporations and large organizations in North America have experienced virus infection of their personal computers. But the good news is that viruses rarely adversely affect computers immediately, and the damage is often minimal. With some basic information, a few precautions, and good software, your organization can use its DIP system with a high degree of confidence that its files are safe from virus damage.

What is a Virus?

A computer virus is simply a program designed to reproduce and spread within a computer system or network without revealing its presence. A computer virus attaches to files or to boot sectors of hard or floppy disks (the portions of the disk that contain system configuration information). Once in a computer system, the virus replicates itself; in fact, some viruses do nothing else. Even so, there is no such thing as a truly harmless virus, since, at the very least, viruses expand, taking up hard-drive space and slowing the performance of a computer.

Viruses are typically classified by the way they "infect" a system. For example, viruses that infect executable programs residing on a computer (such as *.exe* files) are referred to as "program viruses." Viruses that infect the boot section or partition tables on a hard disk or floppy disk are called "boot viruses." Viruses that are capable of infecting both are referred to as "multipartite viruses." Viruses that are contained within automated step sequences in a data file are called "macro viruses," due to the

fact that these step sequences are called macros by many software manufacturers. Because a virus is simply computer code, once it enters a single PC on a system, it can spread by being transmitted along with any legitimate software or file.

Once a computer is infected by a virus, many types and levels of damage can result. Some levels may be considered negligible, as in the cases of viruses that do nothing more than replicate or viruses that alter or delete files that can be easily reloaded when the viruses are removed. Other viruses might destroy all the files on a computer's hard drive, and although this sounds like an overwhelming loss, if backups have been made regularly, even this level of damage can be easily corrected by removing the virus and reloading programs and data from the backups. More difficult to recover from are viruses that slowly corrupt data over a long period of time without ever revealing their presence in the system. In cases such as this, backups are corrupted as well, and data must be rebuilt from scratch. Perhaps the most dangerous types of viruses are those designed to do nothing except discover and report crucial system information, such as the password of the system administrator of a network. If one of these "discovery" viruses is successful, unlimited damage to the system and its data is possible.

Recognizing a Computer Virus

Although these descriptions are chilling to those of us who depend heavily on our computers, there are simple ways to protect your DIP system from these viruses, and ways to recognize when certain types of viruses are residing on the system. Some of the symptoms of a computer infected by a virus include the following:

- Changes in the sizes of programs.
- Longer load times for programs.

- Slower system operation.
- Unusual error messages.
- Reduced memory or disk space.
- Unusual screen activity.
- Incorrect changes in file dates and time information.
- Unexpected writes to a drive.

It is a good idea to share a list of common virus symptoms with every user on your DIP system, then if one PC or file becomes infected, corrective steps can be taken before the entire system is afflicted.

Preventing Virus Infection

There are a few relatively simple steps you can take to eliminate the majority of virus threats to your DIP system. The first, and most important, is to protect against the damage a virus might do by keeping complete and regular backups of your system's files as discussed above. In addition, follow the guidelines below to the minimize the risk of virus infection on your DIP system:

1. Backup all your work and system files regularly.
2. Install and use reputable anti-virus software.
3. Update your anti-virus software regularly.
4. Always scan floppy disks and CD-ROMs for viruses.
5. Do not boot a PC from a floppy disk unless you are sure the disk is virus-free.
6. Whenever possible, use the write-protect tab when using floppy disks to prevent viruses from copying themselves.
7. When downloading from the Internet, scan files for viruses before running them.
8. Use only licensed copies of software obtained from reputable sources.

The most effective way to avoid virus damage is to obtain virus detection software and to use it consistently. Most of these programs work by allowing you to perform scans of random-access memory (RAM), hard disks, and floppy disks for suspicious strings of executable code. Several, such as Symantec Corporation's *Norton AntiVirus*, also allow you to inoculate files on floppy disks or hard disks to detect future virus infection (that is, the deposition of viruses known to the program). Some programs even allow you to keep them running in memory so that an infected program file can be caught as soon as it is copied, launched, or identified. And the newest and most sophisticated virus protection software automatically scans files as they are downloaded from Internet sites to determine their safety.

One of the most important characteristics of virus protection software is its currency. A particular software package can only recognize viruses that have been identified, classified, and defined in the program. For this reason, most of the leading software manufacturers maintain extensive update programs through their Web sites. For example, Symantec Corporation (**http://www.symantec.com**) provides a wealth of information pertaining to *Norton AntiVirus* software.

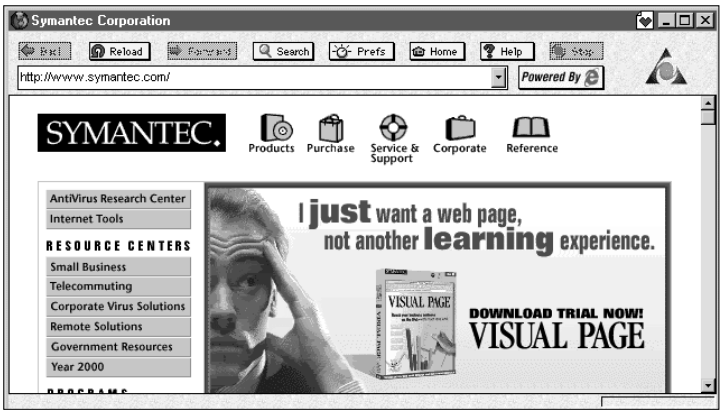


Figure 7.4 *Symantec Corporation's Web site.*

Document Security

Your entire DIP system is built around the goal of creating an electronic archive of documents that can be quickly and easily accessed by the staff members who need them. But these documents will be of very little use if they inaccurately represent the content of originals, if various versions of the documents cannot be distinguished from one another, or if the authenticity of the document is in question. In this section, we will overview some of the points that DIP system managers must consider to ensure that the documents in the DIP archive are secure.

Document Accuracy

When making the transition from manually handled document control processes to automated processes, it is important to establish controls and guidelines that will ensure that the documents that enter your organization's electronic archive are accurate representations of the originals. To do so, managers must identify (1) the ways that data is captured, (2) the processes that are most likely to result in errors, and (3) the procedures that will ensure that errors are identified and corrected before a document enters the archive and the originals are destroyed.

In general, each phase of the document's life cycle in the system must be examined for potential risks and errors. Then, security measures must be put in place to correct those errors. For example, when considering the capturing stages of a DIP system, scanning procedures should be examined closely. How are scans that result in poor quality images identified and enhanced? Are original documents that arrive in the scanning queue in poor shape identified to receive special attention? In an organization in which multi-page documents are scanned, a close analysis of the system may reveal that pages frequently stick together when fed

into the scanner automatically. The result is an incomplete document. Appropriate procedural checks for these kinds of errors might include the following:

- Staff members who coordinate scanning might affix a notation to the first document in a group that indicates the number of pages that should be scanned.
- When the document is scanned, the staff member can check the electronic record of scanned pages against the expected page tally.
- If the appropriate number of pages are scanned, the document can continue on for further processing; but, if the number of scanned pages do not match the document, the electronic file might be dumped and the document queued for rescanning.

Obviously, many other scenarios are possible and likely. The only way to prevent them from compromising the success of your DIP system is through careful analysis and planning.

Similar examples may be identified in each phase of a document's life cycle. For example, it is a rare occurrence for a document to be OCRed with 100% accuracy. After the conversion, most documents must be reviewed by a staff member who corrects conversion errors. Obviously, without this manual check, the accuracy and completeness of the document would be unsatisfactory. Further, documents that have OCRed badly, or images created from damaged documents, may result in incorrect indexing choices. Human intervention is frequently required to identify and correct these types of errors. If these errors go uncorrected, the document may be stored incorrectly and subsequent search-and-retrieval efforts are likely to result in failure and lost documents.

When exploring software products for your DIP system, be certain to evaluate the methods used to resolve these types of issues. For example, the figure below shows the *File Magic Plus* Proofing Editor in which questionable translations of an OCR conversion are identified through a highlighting scheme.

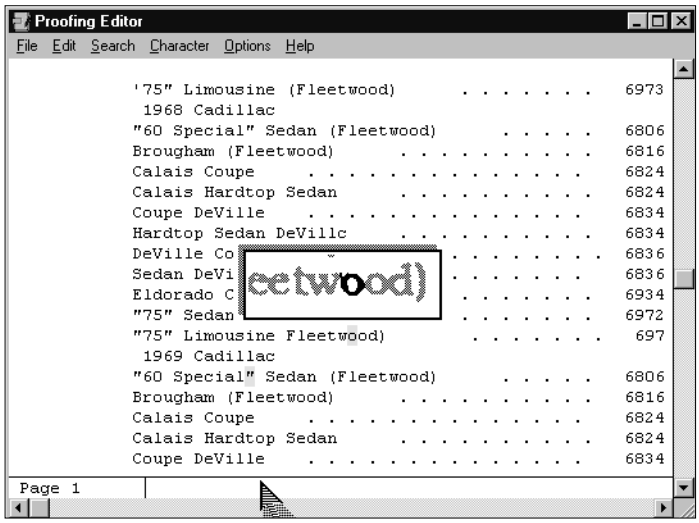


Figure 7.5 *Verifying authenticity after an OCR conversion.*

Document Version Control

One of the primary goals of your DIP system is to make documents easily available to staff members, who can then retrieve them and use them to complete their work tasks. In many cases, the tasks that these staff members complete involve only viewing the documents; but in other cases, documents may need to be annotated, edited, or changed in order for the workflow of the organization to proceed efficiently. In cases like these, it is important that your DIP system is able to supply sufficient version control to ensure the integrity of your document archive.

In part, these version control functions may disallow edits entirely. Documents protected in this way can only be edited by staff members with appropriate security clearance to bypass this

document "lock out." Typically, managers who are responsible for certain types of documents are authorized to restrict their access. Other types of documents may be protected from change from the moment they enter the system–such as certain confidential, high-level contracts that are available for viewing only by high-level managers. On the other hand, if edits are allowed, the version control component of your system will most likely record a version number, the date of the edit, the user who edited the document, and, perhaps, a copy of the original will be stored as well.

Industry leader, PC DOCS (**http://www.pcdocs.com**) has included extensive version control features in their flagship product, *DOCS Open*. This integrated document management suite boasts a library management feature that provides administrators with the ability to store as many as 99 versions of a document, as well as a document history that logs every activity performed on a document during its life cycle. Additionally, *DOCS Open* allows administrators to grant or deny access rights, and the program's Check-in/Check-out feature records documents that have been copied by users, and allows administrators to temporarily lock documents–preventing any alterations to the document for any length of time that the administrator determines.

Of course, your particular version control needs will depend on the types of documents processed by your system and the levels of security restriction that your organization has decided to implement for those various document types. When you have categorized your organization's documents (most likely in your written project plan), you will find it easy to develop a list of version control functions that will be necessary in your system. Use this list when assessing system software options for your organization.

Document Authenticity

In order for electronic documents to truly replace their paper counterparts, your DIP system must ensure that the electronic documents housed in your archive are authentic reproductions of the originals. As we mentioned before, a rigorous system of checks and balances is essential at every step of the DIP process to preserve the integrity of your electronic information. Indeed, according to the AICPA publication, *Audit Implications of Electronic Document Management*, "The risk of fraud–criminal conduct aimed at intentionally altering, destroying, or counterfeiting documents–increases if access controls and management trails are not adequately maintained, monitored and enforced." Obviously, the authenticity of your documents can have far-reaching legal ramifications.

For example, imagine the dire consequences if OppenheimerFunds Services (review this case study in Chapter Two, if necessary) suddenly found their system's document authenticity to be questioned. As the transfer agent for more than 70 multi-class mutual funds, OppenheimerFunds Services must be confident that their document integrity is impeccable. To avoid such possibilities, OppenheimerFunds Services' chose FileNET's system, which incorporated a new customer signature card committal tool. This tool allows the user to capture a digital image of the customer signature card, match it with a check writing control number, and store it. In processing future transactions, the signature card can be displayed very quickly for verification. Read more about OppenheimerFunds Services success with EDM at **http://www.filenet.com/cust/oppenheimer.html**.

So, what steps can you take to ensure the authenticity and integrity of your DIP documents? Again, the best way to proceed is to identify likely risk areas by assessing the DIP system processes. Two areas of risk will be immediately apparent: (1) the

documents themselves and (2) the indexing system that identifies the documents. In EDM systems in which documents from various applications (such as word processors) are stored in their original electronic forms, documents can be changed by simply opening the file in the original application and typing the desired changes. If the system does not use version controls to identify when and by whom documents have been changed, the altered document can simply be saved as a replacement for the original. Even image files can be altered by a user who has limited abilities with a graphics editor. In short, the pixels that make up the document can be altered and the image subsequently changed. If the system on which the image is stored does not lock out or track a change of this kind, then the image can simply be stored in the archive, replacing the original.

Even the inherent security of WORM technology can be bypassed by a clever enough user. For example, although WORM storage prevents a user from overwriting a file that has been stored in an archive, the document can be replaced by altering the index so that it retrieves a different, inauthentic document when a user enters indexing information that should retrieve the real document. These types of indexing problems may be intentional, as in the scenario described above, or they may be accidental, resulting from errors in the assignment of indexing fields and numbers when a document is prepared for archiving. Either scenario has the potential for corporate headaches; but, as the saying goes, forewarned is forearmed–ensuring document authenticity should be an essential element of the DIP planning stage.

Obviously, implementing version control procedures will have an enormous impact on your ability to distinguish authentic originals. In addition, documenting and enforcing appropriate procedures for preparing documents for storage can reduce

accidental indexing and storage errors. Finally, implementing appropriate user access controls will help to protect your system from attacks intended to compromise the authenticity of system documents. The sections that follow will deal with these types of security issues in more depth.

Lastly, the US laws governing the admissibility of DIP documents vary from state to state. So, these issues must be considered when planning your system. And, for those companies that conduct international business, the laws of other countries must be taken into account. In short, there is no specific set of all-encompassing rules that will ensure the legality of your DIP documents. The best course of action is to consult with your legal department or with your own attorney as you plan your DIP system.

User Controls and Access Rights

The third category of risk for which CPAs and financial managers who are responsible for DIP administration and security must prepare is the risks that users, authorized and unauthorized, bring to the system. In this section we will address two strategies for controlling user access: (1) passwords and (2) user groups and permissions.

Passwords

Passwords are the oldest line of computer defense. Although passwords have been used for almost 30 years, they are still frequently misunderstood and often misused. Employees who think nothing of using a key to get into their offices may balk when asked to enter a password to use a DIP system. You, or your system administrator, should ensure that each employee who uses your system also uses a safe and effective password for

individual access. This is important in any networked environment, and it is especially crucial for systems connected to the Internet, where the potential for unfriendly break-ins is much higher.

Some common password problems are easy to avoid. The most common password security problems are called "Joes." Joes are account passwords that are variations of the name of the account's owner. One consulting company theorized that there is at least one "Joe" account on every system, and found at least one such account on every client's system that the team tested. This problem most often results when users are assigned default passwords (such as their last names) when accounts are created, and then employees fail to change the passwords to something more secure. To avoid this problem, accounts should never be given default passwords that are easily guessed, and users should be informed how to change their passwords and should be advised to do so the first time they use their account to log-in to the system.

Another password problem occurs when accounts have no password or some commonly known default password. Some application programs that need access to shared information will create an account for the program at installation and set the password to some common default. To prevent problems like these, always make sure that you understand the installation of your system software and its effect on the password files. If the software requires shared accounts, be certain that you change the password to something other than the default so every purchaser of the software doesn't have access to the account on your system.

Some users of your system may have accounts on other systems. These users will often use the same password for each account

they hold. Using the same password for different systems is secure only when:

- The two systems are logically equivalent, like two file servers in the same company.
- The two systems are run by the same information system center with the same security measures.

If the systems are run by different organizations, have different security levels, or have different operating systems, users should have unique passwords for each system. Let's say that you have an account on your firm's DIP system and also with a commercial database system. If someone cracks your password on the local system and they discover that you have an account on the database server, the cracked password will be the first thing they try when they attempt to gain access to your other account.

You often find warnings about not writing your passwords down. Keeping them online in a text file is certainly not advised, nor is writing them on Post-It notes and sticking the notes to your monitor or to the underside of your keyboard; but keeping a note containing your password in your purse or wallet is probably OK. It is better to have a secure password that you have to write down to remember for the first couple of weeks than it is to have a password that is so obvious that anyone could crack it. You may wish to use the Password Security Checklist on the following page to help your staff members evaluate the security of the passwords that they use on your DIP system.

Password Security Checklist

yes/no

1. Is your password shorter than eight characters? Increasing the length of your password from six to eight characters forces a cracker to check many more possibilities. _____

2. Is your password "letmein," "password," "hello," or some other clever response to the password prompt? These are commonly known and are the first checked by cracker programs. _____

3. Is your password your log-in name in any form? Is it your license plate number, telephone number, street name, or the make of car you drive? Information like this is easy to discover, and so it makes an insecure password. _____

4. Is your password all digits (like "12345678"), or a repeated character (like "aaaaaaaa")? Is it a repeated sequence of digits, or a keyboard pattern (like "12341234" or "asdfasdf")? Passwords like these significantly decrease the length of time needed to crack a password. _____

5. Is your password in <u>any</u> dictionary? Crackers first check words in online dictionaries, both in English and some other languages. _____

6. Is your password a word in a dictionary modified by prepending or appending a digit? Passwords like "house1" are not secure; cracker programs check appended and prepended digits as a matter of course. _____

7. Is your password a word constructed by rotating or reversing one of the above classes? For example, Frank may think that "knarf" (his name spelled backwards) or "rankf" (his name with the first character moved to the end) is more secure than his name, but this is not true. _____

8. Is your password an example taken from a printed source? _____

If you answered “no” to all of the questions above, your password is probably secure. If it is not, consider the following suggestions:

1. If the system distinguishes between upper and lower case letters in a password, mix the case of letters in your password. “TomCaT” is more secure than “tomcat” or “TOMCAT.”
2. Use non-alphabetic characters INSIDE the password. “Tom3CaT” is more secure than “TomCaT” or “TomCaT3.”
3. Choose a line or two from a favorite song or poem and use the first letters of each word. For example, “To be or not to be, that is the question” becomes “2Bon2btitq?.”
4. Choose two short words and concatenate them together with a punctuation character between them. For example: “dog+rain,” “book!mug,” “kid?goat.”
5. Choose an adjective, a noun, a verb, and an adverb. This will make a (possibly nonsense) sentence that you can remember. Now use the first two letters of each word for your password. For example: “Orange Cars Fly Silently” becomes “OrCaFlSi.”

Access Controls

Another aspect of control that you should consider for your DIP system is establishing access controls based on user knowledge needs. One of the great benefits of implementing a DIP system–the ease and speed with which data can be shared among users–can be problematic when some of that data is not meant to be shared. You will probably wish to protect certain types of confidential or proprietary information that resides in your archive. Again, the people who pose a risk for compromising this information reside in two places: inside your organization and outside your organization.

When thinking about internal information access rights, you can use the simple analogy of a filing cabinet to consider the threats and alternatives for protection. For instance, if you are the payroll manager of a large corporation, it is likely that you keep your hard-copy payroll files under lock and key all day long. Probably only one or two other persons have access to these files–maybe your secretary and perhaps your boss, the vice president of administration. If a manager of a particular department comes to you for payroll information on her department's employees, you will release to her the files relevant to her own employees, not the files for her peer managers or the employees of another department. It would be understood that this manager would not release the payroll information for one employee to another, and that she would not publish her department's payroll information company-wide or world-wide.

Now, for comparison, let's suppose that you have an automated payroll system that employs DIP technology to store archived documents related to payroll functions. Perhaps that system has built-in password authorization levels in place for managers to access information in their employee's payroll records. Perhaps the system administrator has implemented a step whereby a

manager can only read that information, but to change it, say, to give an employee a pay raise, she must have the approval of you, the payroll manager. These are useful security measures, and you should use them as is appropriate for your organization, but their efficacy depends on each authorized person using the information only as intended. For example, what good would the multi-level password system be if the manager accessed the confidential information in an authorized manner, but then printed it on her laser printer, photocopied it, and gave a copy to each company employee?

What you need in this instance is a clear definition of the information access rights of each level of employee in your organization and strategies for implementing these controls in the DIP system. One of the most common strategies for contending with this problem (as mentioned in Chapter Six) is identifying the document access needs of certain groups of users and then using DIP software controls to prevent users from accessing any documents that are not identified as necessary to their jobs. Making access decisions requires careful consideration of the tasks performed by various categories of staff members. For example, in a small CPA firm, this may mean that the seasonal tax preparers have full access only to the documents that are relevant to the particular accounts for which they are responsible (that is, they can read, edit, change, and delete them). Or, perhaps, these tax preparers have full access to those documents related to the current tax year for their accounts and read-only access to documents related to their accounts from past years (that is, they can view but not change documents archived from past tax years). The screen below illustrates the Document Access window from which user and group access settings can be selected for a document in *DOCS Open*:

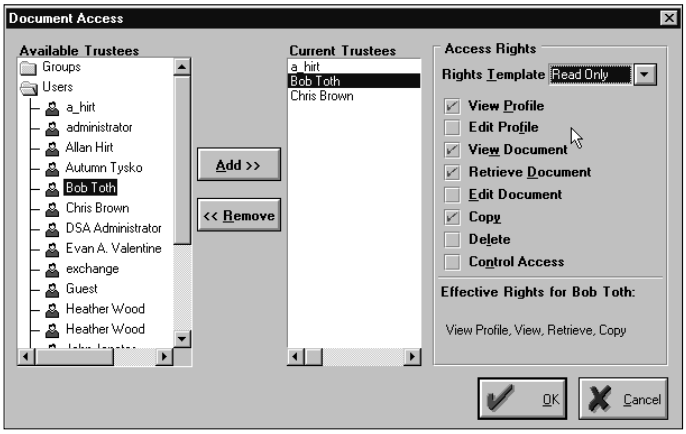


Figure 7.6 *Access control settings in DOCS Open.*

When deciding about information access rights, you can use a rule of thumb common to the medical and legal professions, called the "need to know" rule. According to this rule, only employees who need to know information to perform their jobs should have it. For instance, a manager who evaluates employees' performance and grants merit raises obviously needs to know her employees' salaries, but very few other people do. Everyone in the company probably needs to know the contents of the company's strategic plan, but very few people need to know the competitor analysis on which certain aspects of the plan were based.

As you work through the levels of information that particular classes of employees "need to know," you can use a matrix, or table, to keep track of your decisions. Although the types of rights that can be assigned and how those rights are managed will depend on the particular capabilities of the software package that you choose to implement, you can use a table to organize your decisions and discuss them with appropriate managers and executives in your organization.

Executive Summary

In this chapter, we have identified some of the important areas of risk that CPAs and financial managers involved in DIP projects must be prepared to address. Although the strategies that you will use to implement controls for these various types of risks will depend, in large part, on the software you select for your DIP system, recognizing these key risk areas in advance will help you to identify software capabilities to add to your "must have" list of features. To help you evaluate the risks and control factors of electronic document management systems, you may wish to use the checklist that appears on the final pages of this chapter. Although developed as a checklist for auditors engaged in testing controls in EDM systems, the checklist is a good resource for financial professionals in all levels of DIP administration who must identify risks and protect EDM and DIP systems.

Sample Audit Checklist for EDM System Components

(Source: *Audit Implications of Electronic Document Management,* American Institute of Certified Public Accountants, Inc., 1997. Reprinted with permission.)

Document Preparation Controls yes/no

1. Is document preparation being performed? _____
2. Have personnel been adequately trained on critical information? _____
3. Does a sample of documents taken after document preparation, but before scanning, reveal any of the following:
   - Duplicate documents? _____
   - Missing documents? _____
   - Incorrect associations between master and subordinate documents? _____
   - Inclusion of superfluous information on documents? _____
   - Inclusion of incorrect documents in the batch to be scanned? _____
4. Of the entire document population, are the documents being prepared providing the most timely information to the entity? _____
5. Is the preparation process being ordered according to priority? _____
6. Is there a method used by the client to ensure that all documents scheduled (for example, checkoff lists, batch tickets, signatures) have gone through the preparation process? _____
   If yes, explain: _____
7. Have documents scheduled for processing been authorized? _____

Document Scanning Controls yes/no

1. From an on-line examination of sampled images, do the document quality settings (for example, 200 dots per inch, 300 dots per inch) allow the user to read all information accurately? _____
2. Is there a maintenance contract or service agreement on the scanner? _____
3. Does the maintenance contract include periodic cleaning? _____
4. Is the setting for the scanning threshold appropriate? _____
5. During an observation of the scanning process or interviews with clients, are there any mechanical problems with document feeding, alignment, or document wrinkling? _____
6. Is the resolution power of the scanner such that it can detect critical images located in very close proximity to one another? _____
7. Where are documents stored before the scanning process?
   _____
8. Where are documents stored after the scanning process?
   _____

9. If the documents are stored in a different location after the scanning process, is it adequate? _____
10. Does documentation exist for the retention periods of scanned documents? _____
11. Are retention periods different than those for documents of the same type which will not be scanned? _____
    If yes, explain: _____
12. Does the retention period allow for errors to be detected and corrected on the image before the original paper document is destroyed? _____

Image Cleanup Controls yes/no

1. Is the image cleanup process being performed? _____
2. Are the personnel performing the cleanup familiar with the documents' information? _____
3. Compare a sample of paper documents to their corresponding digital images for the existence of periods, commas, lines, speckling, and other markings on one image only. Are differences apparent? _____
4. Will the markings, or lack thereof, result in a misinterpretation of information? _____
5. Through a comparison of documents, does the existence of marks indicate that the document did not go through the cleanup process? _____
6. Has information been added to the document? _____
7. Is the new information complete, accurate, and approved by appropriate personnel? _____

Index Controls yes/no

1. Are indexes for image files stored on:
   [ ] Optical media? [ ] Magnetic media? [ ] Magneto-optical media?
2. If indexes are stored on magnetic or magneto-optical media, have only appropriate personnel been given security privileges to update the index? _____
3. Is the retention period, destruction date, scanning date, or other information readily accessible from the index that indicates the appropriate time period that the image must remain in the system? _____
4. Is enough information contained in the index to make the accessibility of the image timely and accurate for the users? _____

Processing Controls yes/no

1. Who can gain access to workflow scripts?
   _____
2. Are all individuals on this list authorized? _____

3. What logs, processes, or reports exist to detect and report changes in scripts?

   _____

4. If these changes are authorized, how are the changes reported to the users to prepare them for a change in operations?

   _____

5. Is each script reviewed and approved by a representative of each department/entity involved in the workflow? _____
6. Are images routed to the correct locations? _____
7. Are they routed within the correct time frame? _____
8. At all locations, are appropriate approvals gained before images are routed to the next location? _____
9. Are all document images in one logical folder routed together (master image and all subordinate images)? _____
10. Does an application log exist which tracks activities and locations of each document image? _____
11. At what points (if any) in its life-cycle can the image be altered?

    _____

12. Can the system detect all occurrences of changes? _____
13. Are management staff receiving reports on the efficiency of the workflow process? _____
14. Does evidence reveal that they review them? _____
15. Do they receive timely reports to make decisions improving the efficiency of the workflow? _____
16. Do these reports contain collective and complete information to result in knowledgeable decisions? _____
17. How are exceptions to the script handled?

    _____

18. Do numerous exceptions justify a script to be created especially for them? _____
19. Who can change the system templates?

    _____

20. What processes or logs are maintained to detect or report access or changes to the templates?

    _____

21. What are the current number of exceptions recorded on average per batch (please list maximums and minimums)?

    _____

    _____

22. How many images are incorrectly identified as legitimate patterns? _____
23. From the test documentation, what are the number of exceptions recorded at each interval of changed system settings? _____
24. Are all possible patterns stored in the system? _____

25. Are patterns stored which should not be? _____
26. Are any automated exception-handling methods in place to assist in the correction of exceptions? _____
    If yes, to what extent is the automation used? _____

Output Controls yes/no

1. Who has access to view production images?
   _____
   _____
2. Are these employees different from those who can view paper documents? _____
   If yes, explain: _____
3. Are restrictions placed on when documents can be viewed (such as, after normal business hours)? _____
4. Are controls in place to detect who has viewed, printed and changed an image file? _____
5. Are controls in place to ensure that printed documents are retained until the corresponding images are verified? _____
6. Do output devices reduce the quality of the original image? _____
7. What authorizations are required for access to view or print an image? _____

Storage Controls yes/no

1. Is compression performed? _____
2. Is the compression ratio appropriate, given the compression technique, for the amount of information to be stored on each document? _____
3. Are image files stored on:
   [ ] Optical media? [ ] Magnetic media? [ ] Magneto-optical media?
4. If they are stored on magnetic or magneto-optical media, have only appropriate personnel been given security privileges to update the data? _____
5. Are all changes to image files approved? _____
6. Are archive disks being stored offsite? _____
7. Are disks appropriately labeled and logged? _____
8. Are the on-site and off-site storage facilities secure? _____
9. Who has access to the archive?
   _____
   _____
10. Are there appropriate controls for destruction of documents?
    - Proper authorizations? _____
    - Proper scheduling for routine destruction? _____
    - Special procedures when litigation is present or pending? _____

Chapter Eight
Internet Resources for Document Image Processing

The Internet has changed the ways that savvy financial professionals gather information about how technological advances can prove most beneficial. In short, spending some time visiting Internet sites devoted to recent advances in electronic document management in general and image processing technologies in particular can help you assess the current needs of your firm, as well as plan future growth. In this chapter, we will describe some strategies for searching for useful resources on the Internet. Topics covered include the following:

- Using search engines.
- Tips for successful searching.
- Relevant resources.

Using Search Engines

In the world of information science, running searches for online information is both an exact science and a creative art. Library professionals spend years in training to use specific search-and-retrieval schemes, as well as specialized research tools. They go on to make careers locating and obtaining information for the rest of us, as any visit to a corporate information center will demonstrate. But now that the technology of the Internet has

delivered access to millions of files and databases to each of our desktops, the rest of us need to learn to survive in the world of targeted searches.

When we begin to research a topic in a library, generally, we start with an index. Research on the Internet is no different. The most important tools you will use when researching topics on the Internet are Internet indexes, known as search engines. Simply put, a search engine is an Internet site running special software that catalogs other Internet sites and allows users to search through the catalog using keywords to find matches. There are a number of search engines currently available with varying sizes of catalogs and various methods for running broad and specific searches. Despite their many small differences, most search engines operate in the same manner. Just follow these simple steps:

- ⇒ Access the search engine by typing the site's address in your Web browser.
- ⇒ Once at the site, look for a query box, which allows you to enter your keyword(s).
- ⇒ Type in the keywords that relate to your topic.
- ⇒ Press the **<ENTER>** key or click the **Search** button.
- ⇒ Wait for the results to be displayed on your screen.

Let's take a look at a sample search engine, AltaVista, available at **http://www.altavista.digital.com**. The opening screen of the AltaVista search engine is fairly intuitive to use. The first step is to select from various settings that control the general scope of the search and the language(s) in which documents may appear, and then enter your keyword(s). In general, in AltaVista and in other search engines, multi-word search strings result in more focused matches, so enter two or three words related to your topic.

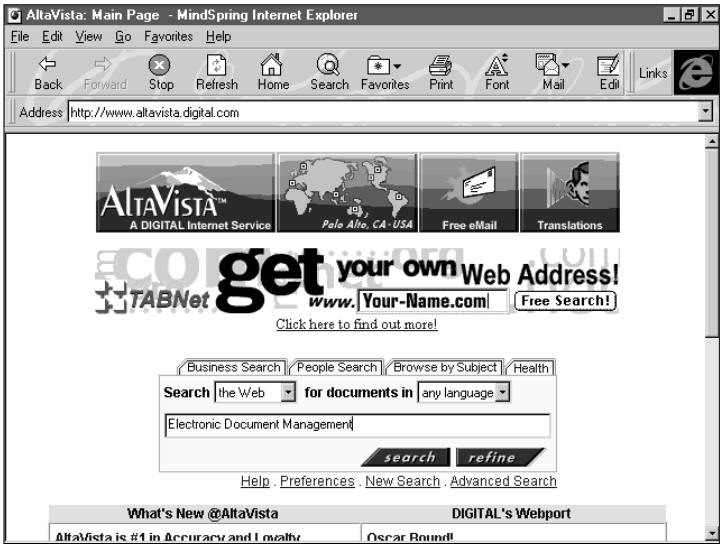


Figure 8.1
Entering keywords in the AltaVista search engine.

When the **search** button is clicked, the program compares the keywords you entered to the items in its database and displays a listing of items that appear to be matches.

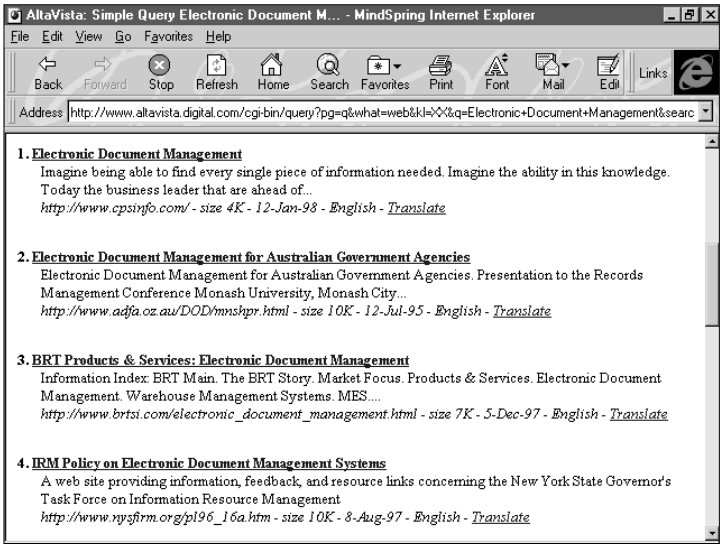


Figure 8.2
An AltaVista "hits" listing.

Each of the listings includes a link to the listed item, so when you find an item that interests you, click the link to access the site. When you are done with the resource, click the **Back** button on your Web browser to return to the listing of hits and select a new lead to follow. It's that simple!

The best way to learn about WWW search engines is to use them and begin to build an experiential sense of what works and what doesn't work. Other sites with WWW search engines and directories that you may want to try are listed below:

ALIWEB	**http://web.nexor.co.uk/public/aliweb/search/doc/form.html**
CUSI	**http://web.nexor.co.uk/susi/cusi.html**
Einet Galaxy	**http://www.einet.net/**
Infoseek	**http://www.infoseek.com**
Metacrawler	**http://www.metacrawler.com**
Lycos	**http://www.lycos.com**
Open Text	**http://www.opentext.com**
WebCrawler	**http://webcrawler.com**
Yahoo	**http://www.yahoo.com**

Refining Your Results

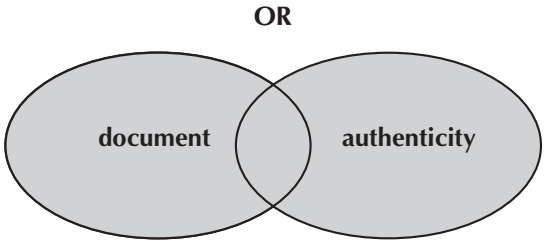
Although you will probably never say, "There is nothing on the Internet about my topic," you may often say, "There are so many Internet resources on my topic, I don't know where to begin!" The simplest way to restrict or expand the number of items you retrieve in a search is to use any of the operators or symbols recognized by the search engine you are using. Although the specific symbols, words, and characters recognized by various search engines may differ, the most common, Boolean operators, are recognized by most. The most often used Boolean operators and the effects they can have on your search results are described below:

OR

The OR operator is useful for the first phases of a search, when you are not exactly sure what information is available on your topic or what words are used to categorize it. When used between two words, the OR operator instructs the search tool to retrieve any record containing either of the words. For instance, the search query

document OR authenticity

would retrieve items containing either the word "document" or the term "authenticity," as illustrated below:



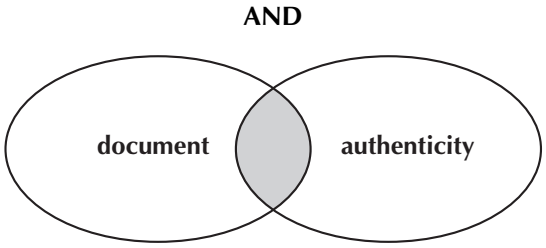
Once you view the types of items containing either word, you might want to narrow your search by dropping one term and confining your search to the other. For instance, you might find that the records indexed under the term "authenticity" are more relevant to your research question than those indexed under "document." Or, as in the example below, you might find that the items related to the specific subject of "document authenticity" must contain both words, not simply either one. Because OR is the Boolean operator that returns the most "hits" (items meeting the search criteria), search queries containing OR are very broad and sometimes return items that are not relevant.

AND

If you need to pose a more specific query, use the Boolean operator AND, which limits results to those items that contain both (or all) of the search terms in your query. Again using the two words from the example above, the search query

document AND authenticity

would retrieve only those items containing both words in the same item, as illustrated below:



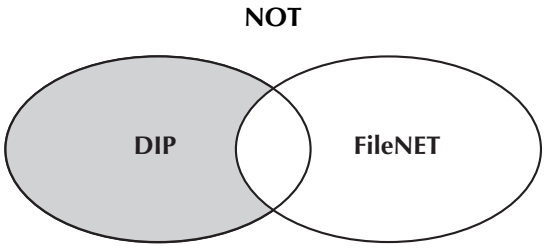
As the shaded area shows, this search query would return a much smaller set of hits, and the items would be more applicable to the topic of document authenticity.

NOT

The last of the three most common Boolean operators is the word NOT. The NOT operator is used to eliminate records containing a particular word or combination of words from your search results. For instance, if you are performing a general search on DIP vendors, you might wish to exclude items dealing with the very specific software vendor, FileNET. To make this exclusion, you could construct your search query as:

DIP NOT FileNET

This search would return all items containing the term "DIP" except for those that also contain the word "FileNET," as illustrated below:



When you visit a search site, always read the instructions or help file before beginning your search. Each search engine has different parameters for using upper- and lower-case letters and combining Boolean operators, and some offer special methods of refining queries. Another good method for refining your search is to run a few searches experimentally to see what results are returned. By browsing through your results list, you can determine whether or not your strategy is returning relevant items. Then, you can construct a search strategy using the Boolean operators OR, AND, and NOT to improve your results. Most search engines assume a predetermined Boolean operator for multi-word search strings. Be sure to consult the search engine's documentation to determine what default Boolean operator is at work.

Other Tools for Restricting a Search

In addition to Boolean operators, there are various other characters and symbols that can be used to refine searches by defining the relationships between words in your search strings. For example, many search engines recognize that words placed within quotation marks should be treated as a phrase; in other words, in order for an item to qualify as a hit, the words in the

search string must appear together. For example, the keyword string **scanners and software** will retrieve any items that contain the word "scanners" and the word "software." But the search string **"scanner software"** will retrieve only items in which the words "scanner" and "software" appear together. As you can imagine, the results of these searches will be quite different.

The best way to determine what symbols and characters are recognized by a search engine is to access the engine's Help or Advanced Searching documents. For example, in the AltaVista WWW search engine that we demonstrated earlier, clicking on the **Help** option opens the document shown below.

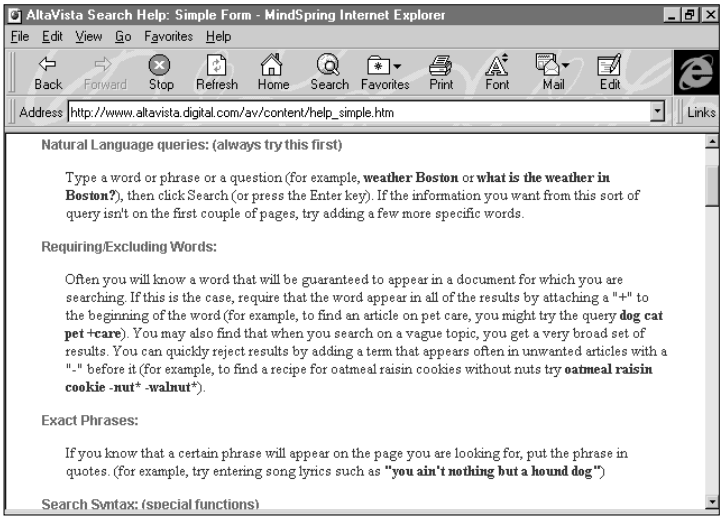


Figure 8.3 *AltaVista's Help document helps you restrict your search.*

Among other things, from this document we learn that AltaVista can support natural language queries (queries that use language that imitates a natural speech pattern rather than specific search terms and characters) and exact phrases. This document also reveals that AltaVista can find pages that include specific graphic files, sort results listings into various topic categories, and much more!

Document Image Processing Resources

Although using search engines will empower you to find myriad useful resources on the Internet, the listing below provides some sites that may be helpful to you as you research possibilities for your DIP project. Review the items listed here and visit any sites that interest you. Please note that although we try to provide readers with current material, resources on the Internet change far too rapidly for any print publication to keep up with. In the event that you cannot access a site listed here, use a search engine to locate the site's new location or to access a site with similar content. To further assist you in your research, remember that many, if not most, sites that offer products and services related to imaging hardware also provide information and links to software as well, and vice versa.

DIP Hardware

(*See DIP Software also.*)

CD Dimensions
This company offers a wide array of products and services in document management services. From scanners to storage systems, this site provides information on all aspects of imaging in a networked environment.

http://www.cddimensions.com

Consan
This company's site offers sales and technical support for a wide range of products, such as Seagate, Quantum, Sony, Pioneer, and Kodak.

http://www.consan.com

Fujitsu Corporation
Another industry giant, this company's site provides in-depth information about all products and services. The site is user-friendly, allowing easy access to myriad relevant and informative links.

http://www.fcpa.com

Hewlett Packard
This site provides an overview of all Hewlett Packard products, new product information, maintenance bulletins, press releases, and Hewlett Packard office phone numbers. The FTP site provides access to drivers, utilities, and support files for all Hewlett Packard products.

http://www.hp.com
ftp://ftp-boi.external.hp.com

IBM Corporation
This industry giant maintains a Web site that is an excellent source for all the products and services that IBM provides. From the home page, you can access information on business solutions, case studies, and sales of specific products.

http://www.ibm.com

Kodak
Imaging industry giant Kodak has created a Web site that is easily navigated. Information on scanners, digital cameras, and printers are but a few of links available on this company's home page. Additionally, case studies, industry white papers, and online technical support are accessed with a click of the mouse.

http://www.kodak.com

Millenial Vision, Inc.
This is a most informative site, particularly for the novice in imaging systems. This site provides a wide range of products and services, as well as information on tutorials and a useful list of definitions.

http://www.mvimvi.com

UMAX Technologies, Inc.
This company is a leader in scanning products and services for any sized business. Like many companies involved in imaging and workflow technology, Umax has created a Web site that allows the user to gather relevant information about the company and access myriad relevant links.

http:www.umax.com

DIP Software

(*See DIP Hardware also.)*

Caere
Caere is a leader in the imaging software industry. This site provides information on current industry trends, as well as information on all Caere products and services, such as OCR and image editing software.

http://www.caere.com

Expert Graphics
The North American distributor of Rasterex products, Expert Graphics' Web site offers contact information, product overviews, free downloadable demos, and product specifications to name a few.

http://www.expertg.com

Filemark Corporation
This site allows you to explore the company's products and services, and provides easy access to information on archival and retrieval software. Additionally, Filemark's site offers case studies, customer success stories, and distribution solutions.

http://www.filemark.com

FileNET, Inc.
Providing information on all their integrated software suites, FileNET's site offers help in assessing the imaging needs of any user. In addition to a wealth of information on products and services, this site offers a free Cost/Benefit Analysis Tool download.

http://www.filenet.com

I.R.I.S. (Image Recognition Integrated Systems)
An international organization, I.R.I.S. manufactures and distributes a family of hardware and software imaging products in both PC and Macintosh platforms. This site provides numerous relevant links, as well as online technical support.

http://www.irisusa.com

ISYS
Developers of a wide assortment of document management, imaging, and search products, The ISYS Web site offers product information, case studies, and free downloadable product demos.

http://www.isysdev.com

Mobius
This site provides a wealth of information on EDM products, services, and solutions; the links to other relevant sites are extensive.

http://www.mobius-inc.com

PC DOCS

At this site, PC DOCS hails its flagship product, *DOCS Open*, as well as its long list of document management systems and add-on products. This site offers customer success stories, system analysis information, and numerous relevant links.

http://www.pcdocs.com

ScanOptics High Speed Document Management Page

This site provides reams of information on products and solutions, as well as links to distributors and resellers.

http://www.scanoptics.com

ScanSoft Inc.

A Xerox company, ScanSoft's site provides a wealth of information that can be especially helpful as you begin your research. Some of the areas the site offers include links to technical bulletins, product registration, a problem report form, and a tip of the week.

http://www.scansoftinc.com

Storage Tools Vendors

Bluebird Systems

http://www.bluebird.com

Cambridge Imaging Technologies

http://www.c-imaging.com

Cheyenne

http://www.cheyenne.com

CyberStorage Systems, Inc.

http://www.cyberstorage.com

Eastman Software

http://www.eastmansoftware.com

Host Interface International

http://www.hostinterface.com/double.html

Identitech
http://www.identitech.com
Innovatech
http://www.docworx.com
INSCI
http://www.insci.com
KOM
http://www.kominc.com
Maximal Systems, Inc.
http://www.maxretriever.com
Micro Design International, Inc.
http://www.indi.com
Optical Imaging Tech
http://www.opticaltech.com
Optika Imaging Systems
http://www.optika.com
Paperclip Imaging
http://www.paperclip.com
Pegasus Technologies
http://www.pegasus-OFS.com
Rising Edge Technologies
http://www.rising-edge.com
Software Architects
http://www.softarch.com
Texas ISA
http://www.texasisa.com
Tracer Technologies
http://www.tracertech.com
US Design Corporation
http://www.usdesign.com

TAC Systems, Incorporated

The TAC Systems Networked Storage Products page provides information on all aspects of optical storage, as well as links to VAR resellers and an extensive online technical support system. Additionally, this site boasts updated downloads for product manuals and a useful FAQ page.

http://www.tacsys.com

Westbrook Technologies

This company is a leader in integrated document management software systems. At their Web site, you can access information on products like *FileMagic*, *FileMagic Plus*, and *CDExpress*, to name a few. Additionally, the site provides links to relevant links and information.

http://www.filemagic.com

General DIP Resources

Associations and Online Magazines

AICPA (American Institute of Certified Public Accountants)
http://www.aicpa.org

AIIM (Association for Information and Image Management International)
http://www.aiim.org

Aslib (Association for Information Management)
http://aslib.co.uk/index.htm

CPA Tech Online
http://cpatech.hbpp.com

Computer Shopper Magazine
http://www.zdnet.com/cshopper

DMA (Document Management Alliance)
http://www.aiim.org/dma/

Data Storage Magazine
http://www.datastorage.com

IHRIM (International Association for Human Resources Information Management)
http://www.ihrim.org
IMC (International Information Management Congress)
http://www.iimc.org
Inform Online Magazine
http://www.aiim.org/inform
Xplor (Electronic Document Systems Association)
http://www.xplor.org/index.html

Case Studies, Profiles, and Demos

FileNET Success Stories
http://www.filenet.com/prods/panagon_info/info_Access.html
http://www.filenet.com/prods/panagon_info/spicer_Corp.html
HMO Gets New 'Image'
http://www2.computerworld.com/home/cwlaunch.nsf/launch?readForm&/home/print9
Kodak Success Stories
http://www.kodak.com/aboutKodak/bu/bis/dmc/successStories.shtml
Opentext Livelink Case Studies
http://www.opentext.com/corp/case-studies.html
Opentext Livelink demo
http://www.opentext.com/livelink/otm_ll_test.html
PC DOCS Customer Profiles
http://208.153.28.8
PC DOCS Product Screencams
http://208.153.28.8/screencams/main.html
Union Bank Case Study
http://www.aiim.org/inform/union.htm

"Eastman Kodak Company Digital Archive: A Strategic Positioning"

This paper, while a promotional document, does provide an excellent overview of digital storage issues and solutions.

http://www.kodak.com/about/Kodak/bu/bis/dmc/whitePapers.shtml

"Estimating the Cost of Scanning: It May Not Be As Easy As You Think"

In this paper, Michael Bida, Product Manager for Scanners and Recognition Products for Eastman Kodak Company, examines some frequently overlooked issues for those who are assessing their scanning needs.

http://www.kodak.com/about/Kodak/bu/bis/dmc/whitePapers.shtml

IMAGE-L

This mailing list is devoted to announcements and discussions related to image processing.

Subscribe through e-mail to listserv@vm3090.ege.edu.tr with the message "subscribe IMAGE-L" in the body of the message.

"Intelligent Data Retrieval from Raster Images of Documents"

Written by Sargur N. Srihari, Stephen W. Lam, Jonathan J. Hull, Rohini K. Srihari, and Venugopal Govindaraju, this paper focuses on document image understanding, OCR, pattern recognition, artificial intelligence, page layout analysis, and document scanning.

http://www.csdl.tamu.edu/dl94/paper/srihari.html

"Knowledge, Distribution and Efficiency"
This article discusses the use of COLD technology for knowledge management.

http://www.aiim.org/inform/Nov97/know1.html

Newsgroups

comp.doc.management

"Reading Between the Lines"
This informative article describes OCR and reviews various OCR products.

http://www.zdnet.com/cshopper/contents/9611/cshp0035.html

Networks (General)

3Com
This leading manufacturer of Network Interface Cards provides technology news, product information, and customer support.

http://www.3com.com/

ARDIS
Claiming to be the largest provider of wireless networking in the US, this company offers information about its products and remote networking capabilities.

http://www.ardis.com

Arthur Andersen's Computer Risk Management
Consulting services for information systems auditing and security.

http://www.arthurandersen.com/bus_info/services/crm/index.htm

Artisoft

This site provides information on products, support and training, and upgrades from the makers of LANtastic.

http://www.artisoft.com/

Bay Networks

This company offers products and services for networks of all sizes, from desktop connectivity to service provider backbones. Specializes in "adaptive networking" for enterprise intranets. The site is searchable by key word and maintains a database about network products, services, technical innovations, and industry trends.

http://www.baynetworks.com

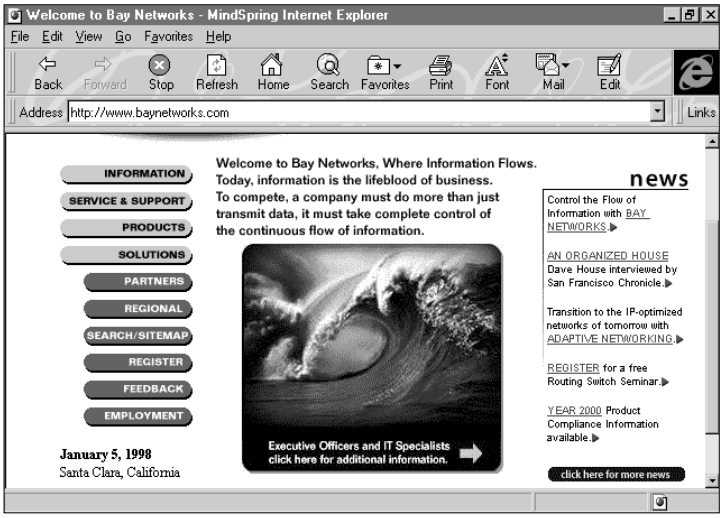


Figure 8.4 *Bay Networks' home page.*

Cabling FAQs

The Massachusetts Institute of Technology hosts this extensive list of Frequently Asked Questions about network cables.

ftp://rtfm.mit.edu/pub/usenet/news.answers/LANs/cabling-faq

Charles Spurgeon's Ethernet Page
This page discusses Ethernet LAN Technology, including the original 10-Mbps system, the 100 Mbps system, and Gigabit Ethernet.

http://wwwhost.ots.utexas.edu/ethernet/ethernet-home.html

Coffeehouse on the Internet
This is a question/answer and comment forum for persons interested in client/server computing. Randy Langel of IBM Southern California Consulting and Services answers daily questions and posts topics that discuss how organizations can implement client/server systems.

http://www.onr.com/oz/house.html

Compatible Systems, Inc.
Provider of Internet solutions for businesses.

http://www.compatible.com/

Data Communications on the Web
This free electronic magazine is designed for managers of corporate networks. The site includes a "Products and Services Buying Guide" that gives results of comparative lab tests of network products.

http://www.data.com/

Digital Equipment Corporation Network Products Page
This section of the company's site includes a product list, announcements of technology seminars, information on training and events, and "customer application stories," which detail real-world solutions. Company stories are indexed by company, industry, and product family.

http://www.networks.digital.com

Dobe Microsystems Online
This firm, located in Eden Prairie, Minnesota, specializes in standard and customized designs for small networks and workgroups. Maintains a "Technology and Trends" magazine and provides lists of readings, a technology news page, and product and consulting information.

http://www.dobemicro.com/

Glossary of Networking Terms
This helpful reference is maintained by Interforce Information, Inc.

http://www.interforce.com/i3spinff/technof/glossary.html

IBM Networking
This page provides links to network hardware products and software solutions developed by IBM. It also features news stories about networking technology, software support information, and numerous links to other networking resources.

http://www.networking.ibm.com

Inside NetWare
This electronic magazine offers tips for users and administrators of *Novell NetWare*. The publisher, Cobb Group, provides a free sample issue online.

http://www.cobb.com/inw/

InternetWeek
Billing itself as "the Home Page for the Networked Generation," this electronic magazine features a news department with audio updates; information about products and services for small networks and the Internet; and NetCentral, a list of online resources about the Internet, intranets, and Local Area Networks.

http://techweb.cmp.com/cw/cwi

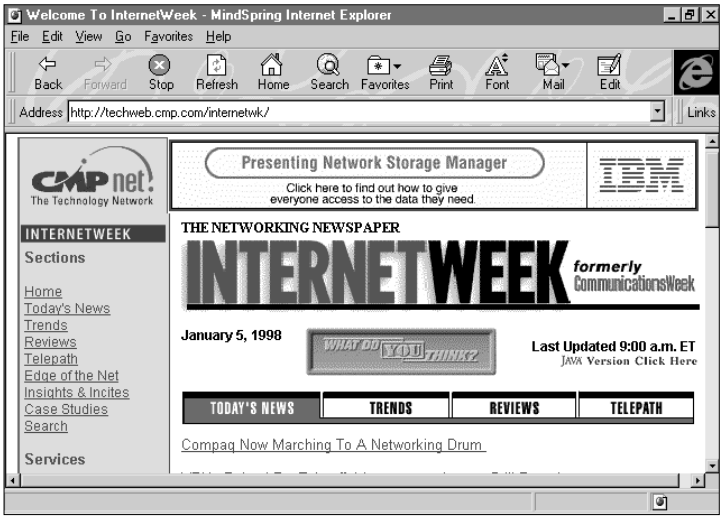


Figure 8.5 *The InternetWeek networking newspaper.*

Intranet Design Magazine

An online magazine devoted to intranet issues and edited by Gordon Benett, author of *Introducing Intranets*.

http://www.innergy.com

Intranet Resources

One of the oldest, most comprehensive online resources about intranets. Features include FAQs, a Discussion Board, Internet and intranet news, articles and white papers, and an "Introduction to Intranets" tutorial for beginners.

http://www.intrack.com/intranet/

Invisible Software

Contains a *Windows 95* troubleshooting guide for installation and implementation of advanced features, such as peer-to-peer networking.

http://www.invisiblesoft.com/

KMJ Communications
Maintains a database of Network Interface Cards organized by type and vendor.

http://www.kmj.com/nic.html

LAN Times Online
This electronic magazine offers news about networking products, feature articles on management and installation issues, analysis of industry trends, and a Testing Center that reviews products. The index is searchable by subject.

http://www.wcmh.com/lantimes

NetWare Connection Online
Aimed toward users of *Novell NetWare*, this site features articles, archives of past issues, and an online bookstore. A hard copy subscription is also available.

http://www.novell.com/nwc/

NetWare Users International
An independent, non-profit group of *NetWare* professionals maintains this site to support *NetWare* group users worldwide. Members receive product demonstrations, conference information, and a monthly newsletter.

http://www.novell.com/nui/

Network Computing Online
This electronic magazine explores all aspects of networking, from hardware and software to management issues. Includes news, features, and how-to articles about product implementation. The subscription is free.

http://techweb.cmp.com/nc/docs/

Network General
A provider of network fault and performance management solutions.

http://www.ngc.com

Network Interface Card Buyers Directory
LANTimes Online hosts this directory of NICs, which includes a product index and direct links to vendors.

http://www.lantimes.com/lantimes/buyers/index/s111.html

Network Management Basics
This document, maintained by Cisco Systems, covers the background and evolution of networking and network management, including issues such as security and network auditing.

http://cio.cisco.com/univercd/data/doc/cintrnet/ito/55018.htm

Network World Fusion
A free electronic magazine about enterprise computing. Maintains a large list of other online resources.

http://www.nwfusion.com/

Networking: A Primer
This white paper at the Bay Networks site describes the evolution of Local Area Networks and explains basic terminology and operations, including how to connect a LAN to the Internet.

http://www.baynetworks.com/Products/Papers/wp-primer.html

Newsgroups

bit.listserv.novell
comp.dcom.lans.novell
comp.os.netware.announce
comp.os.netware.connectivity
comp.os.netware.misc
comp.os.netware.security
comp.dcom.lans.ethernet
comp.dcom.lans.misc
comp.dcom.lans.token-ring
comp.dcom.net-management
comp.protocols.tcp-ip

Novell

The company maintains a large database about its products and services. The home page links to separate departments, including Intranetware, GroupWise, ManageWise, Beta Programs, and Novell/NT integration.

http://www.novell.com/

Novell NetWare Self-Teaching Tutorials

A list of video tutorials that help users master new and advanced features of *Novell NetWare*. The tutorials are priced at $59.95 each.

http://www.candico.com/training/windows/n/netwarecne4.html

Novell Software Utilities

Maintained by Softseek.com, this site lists utilities and tools for *Novell NetWare*, including auditing, e-mail, and security helper applications. Click on the product name to read a review.

http://www.softseek.com/Utilities/Networking/Novell/index.htm

Small Business Networking Solutions
Sponsored by Joseph Williams and Associates, an independent consulting firm, this site offers overviews of *Windows 95* peer-to-peer networking and *Novell* systems.

http://jwa.perdido-key.com/network.htm

Virtual Motion
This company develops and markets software products for LAN-to-Internet access, network security, remote networking, and network management.

http://www.virtualmotion.com/

Windows 95.com
A service of Jenesys, LLC, this site maintains a database about *Windows 95*, covering user tips, 32-bit shareware and driver updates, and bugs and fixes. The Networking department features advice on installation and troubleshooting, and an Internet glossary.

http://www.windows95.com/

Security

Building and Auditing a Trusted Network Environment with Netware 4
Novell has included a great deal of general information about auditing networks at this site. They also market their own network software and auditing software.

http://occam.sjf.novell.com/nw410.english/trustenu/1.toc

CERT
Computer Emergency Response Team's server with articles about security concerns, tools for evaluating security, and an archive of alerts about break-in attempts.

ftp://cert.sei.smu.edu/pub

CERT Security Advisories
Indexed list of Computer Emergency Response Team's warnings issued to mailing lists.

WAIS cert-advisories.src

Check Point Software Technologies
Producer of network security solutions, including firewalls.

http://www.checkpoint.com/

CMS Technologies
A provider of physical security products for networks and stand-alone PCs.

http://www.cmstech.com/

Computer Security Resource Clearinghouse
Maintains security awareness and training information, publications, conferences, software tools, as well as security alerts and prevention measures.

ftp://csrc.ncsl.nist.gov
gopher://csrc.ncsl.nist.gov
http://csrc.ncsl.nist.gov/

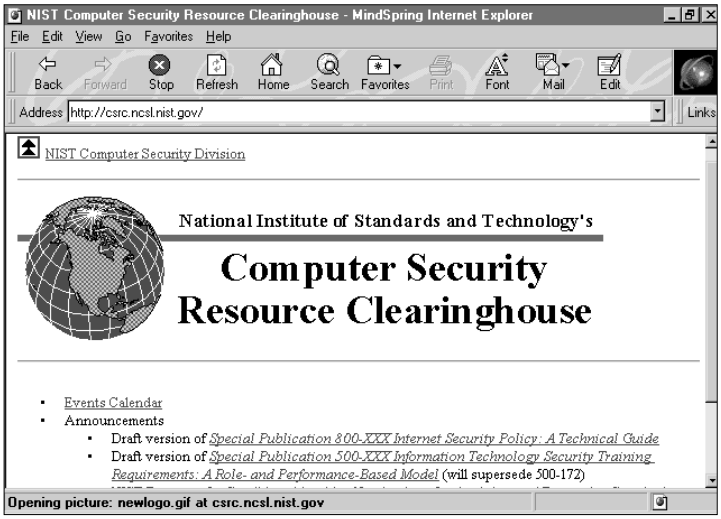


Figure 8.6 *Computer Security Resource Clearinghouse home page.*

The Computer Security Institute
This site provides links and information about computer and network security, including a series of electronic manuals for managers.

http://www.gocsi.com

Computers and Information Resources (CIRT) Home Page
This site provides links to resources about ethics and security. The site was established by Dr. David Grisham, Security Administrator for the University of New Mexico.

http://www.unm.edu/cirt/irc/svcs/secpage.html
ftp://ftp.unm.edu

Cylink Corporation
Specializes in encryption and network security solutions.

http://www.cylink.com/

Data Fellows Virus News Updates
At this site, you can access a database of information about computer viruses either alphabetically or through a keyword search.

http://www.datafellows.fi/news/vir-news/

Dr. Solomon's Online
An information center for users of Dr. Solomon's anti-virus software, this site also provides a "Virus Encyclopedia" and articles on viruses and hoaxes.

http://www.drsolomon.com

Dynasoft
This company's products include computer security solutions for client/server environments, including *UNIX*, and a smartcard-based PC security system.

http://www.dynas.se/

Firewalls mailing list
An e-mail discussion group focused on the subject of firewalls and Internet security.

Subscribe through e-mail to majordomo@greatcircle.com with message: subscribe firewalls-digest

GreatCircle Associates
A firm that specializes in training about and consulting on Internet security firewall systems.

http://www.greatcircle.com/

A Guide to Understanding Audit in Trusted Systems
This useful resource for auditors is presented by the National Computer Security Center and offers guidelines for Internet and intranet security audits.

http://bilbo.isu.edu/security/isl/audit.html

IBM Anti-Virus Online
An electronic magazine that disseminates information about computer viruses through articles, virus and hoax alerts, and technical support from IBM labs.

http://www.av.ibm.com/

Information Security Discussion List
A non-moderated Internet discussion list for information security and auditing professionals in government, industry, and academic institutions.

Subscribe through e-mail to listserv@etsuadmn.etsu.edu with message: SUB INFSEC-S yourname

InterNIC Site Security Handbook
Currently under revision, this document is a popular reference guide for Internet security issues.

http://internic.net/nic-support/fyi/rfc2196.html

McAfee Software

Maker of the anti-virus software, *VirusScan*.

http://www.mcafee.com

Memco Software

Provides enterprise security solutions, including *UNIX*, single sign-on, and distributed security administration. The site also offers security news and resources.

http://www.memco.com/

National Computer Security Association

An independent organization that acts as a clearinghouse on security issues affecting networks and stand-alone PCs. The site offers information on cryptography, firewalls, and anti-virus products, and hosts a message board for the discussion of security topics.

http://www.ncsa.com

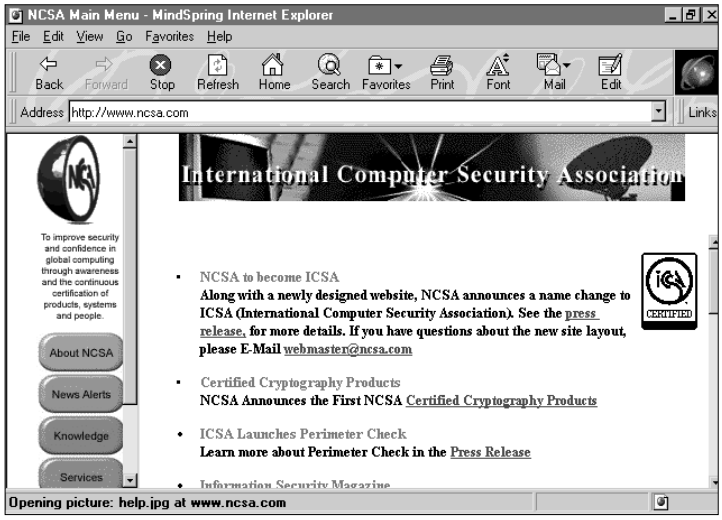


Figure 8.7 *National Computer Security Association home page.*

Netscape Data Security
Netscape Communications devotes a section of its site to discussing Internet security concerns and how its SSL protocol works.

http://home.mcom.com/newsref/ref/netscape-security.html

RSA Data Security, Inc.
This network and computer security software vendor publishes white papers as well as a FAQ on cryptography export laws, which can be downloaded and read in *Adobe Acrobat* PDF format.

http://www.rsa.com

Security Dynamics
Provides network security hardware and software products. Holds a patent on the SecurID Card, a personal identification token supported by various host, server, and network systems.

http://www.securid.com

Security First Technologies
Develops system security for *UNIX*, networks, and Internet communication authentication products.

http://www.sware.com/

Snake Oil FAQ
This document by Matt Curtin, chief scientist with Megasoft, Inc., explains how encryption systems work and offers tips on how to choose encryption software.

ftp://rtfm.mit.edu/pub/usenet/news.answers/cryptography-faq/snake-oil

Symantec Antivirus Research Center
Symantec, maker of anti-virus software, maintains a research library about computer viruses and Symantec products, and gives special coverage of *Macintosh* viruses.

http://www.symantec.com/avcenter/index.html

Technologic
An Atlanta-based corporation that specializes in network security and firewalls.

http://www.tlogic.com/

Trusted Information Systems, Inc.
Provides security analysis and risk assessments, firewalls, mail encryption, and other tools for network and computer systems security.

http://www.tis.com

The World Wide Web Security FAQ
A frequently-browsed FAQ providing an overview of security concerns and solutions. Topics include: threats to security; how to run a secure server; how to protect confidential documents; CGI scripts; client-side security; and information on *Windows*, *UNIX*, and *Macintosh* servers.

http://www.genome.wi.mit.edu/WWW/faqs/
www-security-faq.html

Executive Summary

In order to retain a competitive edge, today's financial professional must be adept in many areas. One of the most important strategies that financial professionals can use to stay current on the technologies that are crucial in today's business environment is Internet research. Understanding how and where to find the most relevant and useful resources can help you solve the problems that you may encounter as your organization moves toward the future. These skills will help prepare you to assess your (and your clients') Document Image Processing needs using the vast resources available on the Internet, 24 hours a day, seven days a week.

ISBN 1-892855-05-4

www.aicpa.org

090434