Electronic Document Management

A ScanCenters' Primer for Real People

by Scott Fordin

for ScanCenters of America

Copyright

ScanCenters of America, National Headquarters
5 Strathmore Road
Natick, MA 01760, USA
Telephone: (508) 655–8000
Fax: (508) 655–2411

THIS PUBLICATION IS PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS PUBLICATION COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY MADE TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS PUBLICATION. SCANCENTERS OF AMERICA MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THIS PUBLICATION AT ANY TIME.

March, 1995

Printed in the United States of America.

Contents

- ✔ You want to know the facts about electronic document management.
- ✔ You want to know how recent advances in computer technology can help you make your business run more efficiently, more error-free.
- ✔ You want to do more for less money, in less time, and with the kind of results that make customers smile—and competitors wince…
- ✔ You want to understand what that hotshot *kid* is talking about when he starts throwing around terms like *scanner*, *workflow*, and *OCR*.

This booklet explains all this and more. Like how you can:

- ☞ Automate much of your paperwork.
- ☞ Save costs on forms management and sign-off procedures.
- ☞ Minimize redundancy and maximize efficiency.
- ☞ Reduce the mistakes associated with manual data entry and storage.
- ☞ Find *quickly* the information you need quickly.

ScanCenters of America has created this booklet so you can decide for yourself whether electronic document management can help your business. We believe the more you know about electronic document management, the more you will appreciate the services provided by ScanCenters of America.

Who Should Read This?

✔ People whose businesses require precise cataloguing and fast access to client, billing, and research data. For example, lawyers and doctors.

✔ People who want to archive their paper records in a more manageable and useful way. For example, accountants.

✔ People whose businesses involve processing large numbers of paper forms. For example, insurance agents, bankers, and government office managers.

✔ People with rapidly evolving product literature or inventory tracking procedures. For example, retail or wholesale sales executives.

✔ People with microfiche or microfilm records they would like to convert to a format that can be used by computers. For example, librarians.

Why *You* Should Read This

☞ Electronic document management is not just where we are headed in the *future*, it is here *right now*.

☞ Businesses today are *re-engineering* and *rightsizing*; effective sharing of information among employees is the key to flattening the decision-making hierarchy, allowing for a leaner, more responsive company.

☞ The biggest difference between people who can take advantage of the latest document management technology and those who can't is not money—it's *knowledge*.

☞ By arming yourself with knowledge, you can make the best decisions for you and your business—technology decisions that *will* have ramifications for at least the next five years to a decade.

How This Booklet Is Organized

This booklet is organized in a way that makes it easy for you to find just the information you want, without having to read the entire booklet from front to back.

- **Section 1, "Electronic Document Management Basics,"** explains the basic terminology and concepts around electronic document management.

  This section includes explanations of indexing, workflow, scanners, storage devices, optical character recognition, and file formats, and how these components work together in an electronic document management system.

  If you want to understand how electronic document management works, read this section. If you don't care about how it works, and you just want to learn about the first steps to take to implement a document management system, read the next section.

- **Section 2, "Taking the Plunge,"** explains the pros and cons of electronic document management (as well as *non*-electronic document management).

  This section also describes how to get started with electronic document management in your business.

  Read this section to find out how or if electronic document management can work for your business.

- **Section 3, "Success Stories,"** describes how three companies have "taken the plunge" and successfully implemented document management systems.

  Read this section to get ideas, inspiration, and encouragement in taking the first steps towards streamlining *your* business with electronic document management.

- This booklet concludes with a handy **Glossary** of relevant terms, and an **Index**.

Electronic Document Management Basics

This section explains the basic concepts and terminology behind electronic document management. Specifically, this section includes the following topics:

The goal of this section is to provide you with an understanding of the components comprising an electronic document management system.

Once you understand the concepts in this section, you will have a clearer picture of what is involved in incorporating electronic document management technologies into your business.

After reading this section, continue on to Section 2, “Taking the Plunge,” on page 25, for information about the costs and benefits of an electronic document management system.

What Is Electronic Document Management?

Electronic document management refers to the process of storing documents—forms, articles, records, correspondence, photographs, and so forth—on magnetic or optical media (see page 3) that can be manipulated by computers.

Such a system requires that the document to be stored either:

- ✔ Already exists in a computer-compatible electronic format

  or

- ✔ Gets converted into a computer-compatible electronic format, usually by means of a *scanner* (see page 8)

Once a document is stored in such a system, electronic document management refers to the various processes related to document search and retrieval, indexing (see page 22), workflow management (see page 23), and distribution.

For example, consider a scenario in which an invoice is sent to your company. At the most basic level, electronic document technology lets you store the invoice in your computer, where it can be easily retrieved for later reference.

Suppose, now, that your company has established rules regarding the routing of invoices above or below a certain dollar amount. Perhaps the invoice is passed directly to accounts payable if the amount is under $500, but gets routed through one or more executive-level approvals if it is above $500.

Electronic document management technology can automatically note the dollar amount on an invoice and route it to the appropriate departments or decision-makers. When all approvals have been obtained, the system can then automatically prepare a check voucher.

What Is a Storage Device?

The underpinning of any electronic document management system is the set of devices on which the documents in that system are stored.

The most common types of storage devices used with computers can be divided into two general categories:

- ☞ *Magnetic media*
- ☞ *Optical media*

Magnetic Media

Magnetic media refers to computer hard disks, floppy diskettes, and digital tape. Like audio or video tape, such media are manufactured by adhering a thin substrate of a magnetically charged substance—commonly chromium dioxide (CrO_2)—to a vinyl, acetate, or metal surface. This surface is then housed in some sort of enclosure or assembly that can be mechanically handled by computer equipment.

For example, a computer *floppy diskette* is an acetate disk enclosed in a plastic case, which in turn is inserted into a diskette drive in a computer. A *hard* (or *fixed*) disk is a metal platter, hermetically sealed and permanently mounted inside the computer. *Digital tape* looks and works similarly to audio or video cassettes.

Data, recorded as changes in the magnetic fields in various regions on the substrate material, is read by computer equipment via mechanical sensors (*heads*), similar to the process used with audio and video tape players.

Before using a magnetic disk for the first time, it must be *formatted*, which is the process by which a magnetic “roadmap” of magnetic fields is created on the disk. The computer then keeps track of the data bits, which are deciphered as changes to this basic roadmap.

Optical Media

The most common types of *optical media* are CD-ROM (like audio CDs), Rewritable, and Write-Once-Read-Many (WORM) disks. As the name implies, optical media are not magnetic at all, but rely instead on the optical reflective properties of the media surface.

Specifically, microscopic bumps, pits, or tiny holes (depending on the type of disk) are etched in the surface of the otherwise smooth media, usually made out of nylon-reinforced glass, covered with a protective plastic layer. These bumps or pits reflect light differently than the rest of the mirror-like surface.

By bouncing a tightly focused laser beam off the media surface and using a photoelectric sensor to measure the amount of light reflected back, it is possible to decipher the pattern of pits or bumps, which are encoded as data on the disk.

From the standpoint of the user, CD-ROM, Rewritable, and WORM disks are very similar—they differ only in the way they are created and/or reused. Specifically:

- CD-ROM disks cannot be re-recorded—that is, once recorded, their data cannot be altered or erased—but you can *add* data to CD-ROMs that have been formatted for *multi-session* use.

- Rewritable disks are able to be erased and re-recorded. Such disks often employ *magneto-optical* technology, which is a combination of magnetic and optical methods, whereby the optical properties of magnetic fields are measured.

- WORM disks are similar to CD-ROMs in that their data cannot be altered or re-recorded. Unlike CD-ROM disks, data on WORM disks can be erased entirely—updating means destroying the existing data and writing new data to an unused part of the disk.

Choosing the Right Storage Device

Just as you want to store your paper documents in a sturdy file cabinet—and your really valuable documents in a safe—the storage devices you choose for your electronic document management system must be:

- ✔ Convenient
- ✔ Secure
- ✔ Safe
- ✔ Cost-effective
- ✔ The right size for your needs

Fortunately, computer-based storage devices outperform their paper-based counterparts on all points. These advantages are explained in more detail below.

Convenience

A standard 3.5 inch floppy diskette can hold 1.44 *mega*bytes of data (1.44 *million* bytes). That translates to roughly 1 million alphanumeric characters (letters, numbers), or 200,000 words, or 800 pages of ASCII-encoded text. And in the world of computers, floppy diskettes are *small*.

A standard CD-ROM disk can hold about 600 megabytes of data, and it is not uncommon to find fixed disk drives with a capacity of 1 *billion* bytes (a *giga*byte)!

As you may imagine, physically handling a stack of 800 pieces of paper is quite different from slipping a floppy diskette into your shirt pocket or putting a removable hard disk into your briefcase.

Moreover, if you want to actually find something in that stack of paper—a specific, *nit-picking* piece or combination of information—it is several orders of magnitude faster and more accurate to let a computer do your sifting.

For example, suppose you are a store owner who wants to find all customers between the ages of 35 and 45, who have made purchases in the last six months, and who live in New York, New Jersey, or Connecticut. You could find that information on a computer disk in a matter of *seconds*, compared to hours or weeks of sorting through papers.

Security

Computer disks, being so portable, can be locked away in safes, or moved off-site to secure locations. Perhaps more importantly, the data on computer disks can be *encrypted*—that is, converted into any number of cryptographic codes that can only be deciphered by someone with the correct code keys or passwords.

In this context, magnetic media in particular provides the following advantages:

- ✔ Easiest to update or modify
- ✔ Sensitive data can be destroyed (digital shredding)
- ✔ Encryption schemes can be changed as often as needed

Safety

Again, due to their portability, computer disks can easily be placed in fireproof safes or vaults. Multiple copies of your valuable information can be conveniently maintained in one or more physical locations.

Optical media in particular cannot be erased or altered without destroying the disk (data on a CD-ROM disk is a legal form of documentation under U.S. law). Optical media is also resistant to flood damage, soot, magnetic disturbances, light, and poor air quality.

Cost-Effectiveness

With convenience and security comes cost-effectiveness. Specifically, if you can find information more quickly and accurately, enter and store it more easily and securely, and do it all on computer-based media that costs about the same as paper, it makes sense that the computer media will be most cost-effective.

- ✔ Magnetic media costs less to create than optical media, and can be re-used.
- ✔ Optical media lasts longer and can hold more data.
- ✔ Digital tape is the least expensive medium for mass storage.

Flexible Capacity

Computer storage devices can easily be expanded to fit your needs. For example, if you need to store another 800 pages, get another floppy diskette or another tape cartridge. When you no longer need to store those 800 pages, use the diskette or tape for something else.

If you need to distribute your company's corporate policy in a reliable way to all your field personnel, create a CD-ROM disk. If you want to store photographs in a digital, indexable format, use a rewritable optical disk.

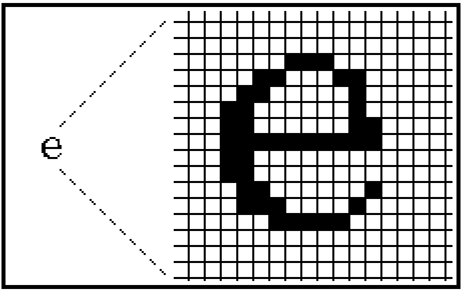
In short, computer media capacity can be expanded or shrunk according to your requirements quickly, easily, and *when you need to*. You no longer need to stack cardboard boxes, stockpile filing cabinets, or pay someone to store or shred your paper records.

What Is a Scanner?

The most common way to convert documents that are not already in a computer-compatible electronic format into such a format is by using a device called a *scanner*.

A scanner is an opto-mechanical device that takes an electronic “snapshot” of a printed page or photograph. Similar to a photocopy machine, a scanner uses a laser beam to bounce light off a page, and then records the amount of light that is reflected back.

Unlike a photocopier, this reflected page image is not transferred to printed (copied) page. Instead, it is converted into a series of data bits, with each bit representing, at the most basic level, either a black or white dot.



This is accomplished by means of a *charge coupled device* (CCD), which is a dime-sized photosensitive chip, mounted on a mechanical arm. The arm is moved across the page in discrete steps; at each step, the CCD measures the amount of light reflected off the page. If the reflected light level is low—below a certain threshold value—it is recorded as a black dot; if high, it is recorded as a white dot.

Just as photographs in a newspaper are made up of very small printed dots, a scanner “prints” a bit pattern representation (called a *bitmap*) of the page. Not surprisingly, the smaller and more numerous the dots that are used to represent the page, the higher (sharper) the *resolution* of the page image.

Scanner Resolution

Scanners commonly operate at resolutions of 75 to 600 dots per inch (dpi). The kind of scanners used in magazine production or advertising may go has high as 2400 dots per inch. In addition, many scanners can record shades of gray (*grayscale*), instead of just black and white, and some can also record color.

- Grayscale scanners work basically the same way as black-and-white scanners. However, with grayscale scanners, each dot on the page is stored as 1 *byte* (8 bits), rather than 1 bit, which allows for 2^8, or 256 shades of gray.
- Color scanners store at least 3 bytes of data for each dot on the page; one each for 256 values of red, green, and blue (RGB). This is accomplished by filtering the scanner’s laser beam with the appropriate RGB color filter, and making three passes over the page with the CCD. Similar to your television, the RGB values are combined to produce a particular color—three bytes of color data can represent 256^3, or 16,777,216 colors!

Types of Scanners

As indicated above, scanners can divided into the following three categories:

- *Black and white*
- *Grayscale*
- *Color*

In addition, scanners come in four basic styles:

- *Hand-held* – Small devices that, as the name implies, fit in your hand. With such scanners, the CCD is not mounted on a mechanical arm; rather it is fixed in position in the scanner housing. *Your* arm replaces the mechanical arm as you pass the scanner manually over the page.

Hand-held scanners are generally least expensive, but are only suited to low-volume applications, or those in which extreme accuracy is not a concern.

☞ *Flatbed* – Most similar in appearance to a small photocopier. Like a photocopier, the page to be scanned is placed on a flat piece of glass, under which the CCD assembly passes.

Flatbed scanners range widely in price and quality. In general, they are suited for high volume environments, like those found in most offices. Many flatbed scanners can be equipped with an *automatic document feeder*, which allows you to load stacks of pages to be scanned, rather than placing them one at a time on the scanner glass.

☞ *Bar Code* – Small, pencil-like wands, flashlight-shaped devices, or fixed lenses beneath glass plates embedded in counter-tops; designed solely to scan bar codes. For example, the scanners found at most grocery stores, which read the UPC price labels on packaged goods.

Bar code scanners are designed to read only bar codes—a specialized function that they do better than any other type of scanner. If you need to scan bar codes, you need to get a bar code scanner.

☞ *Drum* – Drum scanners are drum-shaped devices, smaller than flatbed scanners, used for scanning photographic prints or slides at extremely high resolutions and color depths.

Drum scanners are usually quite expensive—starting at around $7,000 and costing up into the tens of thousands. Such scanners are used by professionals who require the highest accuracy and color fidelity for publication work in magazines, advertising, and the like.

Data Storage Considerations

With higher resolutions, grayscale data, or color depth comes increased storage requirements. For example, at 300 dots per inch, there are 8,415,000 dots on an 8½ by 11 inch page. This translates to 1,051,875 bytes for black and white; 8,415,000 bytes for 256 shades of gray; and 25,245,000 bytes for color!

This kind of storage consumption can add up quickly. To remedy the problem, you need to strike a balance between the kind of scanning you want to do and your storage capabilities. Specifically:

- ✔ If you are scanning text—which is what most people do—stick to black and white. Little advantage is gained with grayscale or color when scanning text.

  In some cases, you may want to convert the bitmap images of your pages into editable text by means of an *optical character recognition* program (see page 13). This has the side effect of making smaller data files—the tradeoff, however, is that performing OCR on a document can be much more labor-intensive than simply scanning it.

- ✔ If you are scanning images (photographs, for example) rather than text, consider carefully whether you really need color. For example, the laser printers found in most offices print only in black and white —all color information is lost.

  Also realize that most line drawings—architectural plans and mechanical drawings, for instance—are black and white. You would do better scanning at higher resolution than worrying about grayscale or color.

- ✔ If you absolutely must have color, make sure you have plenty of storage space, and use the best data compression scheme you can find (see below).

Data Compression

Data is stored in named groupings of bytes referred to as *files*. For example, if you scanned a photograph of your house, you could store the bitmap image of that photograph on a disk in a file named `myhouse.tif` (more about that `.tif` part of the name later). *Data compression*, then, refers to the process of reducing the size of a file without losing any of the data it contains—or at losing only unnecessary data. *Decompression* refers to the process of restoring the file to its original state.

Data compression can be likened to freeze-drying packages of food (as used by hikers and astronauts). By removing most of the water from the food, the food packages can be made lighter and smaller. Decompression, then, is analogous to reconstituting the food—putting the water back in—restoring it to its original condition.

For example, if you used a grayscale scanner to scan a black-and-white image, you could save an enormous amount of disk space by throwing away any data that relates to gray tones, and translating everything to either black or white.

There are numerous data compression methods, each based on a complex set of mathematical algorithms—rules, tools, and assumptions encoded in a software program or hardware chip. For example, MPEG-1 (Motion Pictures Experts Group) compression is used to compress and uncompress video and sound.

The data compression method that is best for you depends on the types of files you are using. Text files benefit from different algorithms than image files, which benefit from different algorithms than sound.

What Is Optical Character Recognition?

Optical character recognition (OCR) refers to the process of converting a bitmap image of a scanned page into ASCII characters. The result of such conversion is a file that you can view, modify, store, or print in a word-processing, spreadsheet, or database program.

Basically, optical character recognition software works by determining the boundaries of clusters of dark bits on a page, then trying to match the shapes of those clusters to alphanumeric characters that exist in its software database. The software then tries to match groupings of found characters with words in its database.

For example, the software may find a shape that looks like a lowercase ‘i’, followed by a lowercase ‘t’. ‘i’ and ‘t’ are both legitimate letters, so the software proceeds to group the two letters into a word. Again, ‘it’ is a legitimate word in the software’s database. If, however, the software found the letters ‘i’ and ‘p’, and then tried to form the word ‘ip’, the questionable word would be flagged by the software as a possible error.

OCR software can also be configured to recognize complex page column layouts and margins, fancy fonts, and text treatments such as underline, bold, and italics. You can then export all this information to your favorite word-processing program, and most or all of the original page formatting will remain intact.

Intelligent Character Recognition

A related technology, called *Intelligent Character Recognition* (ICR), matches letter shapes to mathematical algorithms representing general categories of letters, rather than specific fonts in specific point sizes like OCR. This lets ICR handle a wider range of documents and type styles than OCR. In addition, ICR uses more sophisticated lexical context rules to determine not only if a word exists in a dictionary, but if pairings of words seem to go together logically or grammatically.

For example, ICR software would recognize a string of letters like ‘quickshot’ as ‘quick shot’, rather than ‘quicks hot’. You can also *teach* ICR software your own custom lexical rules, so if ‘quicks hot’ is an expression you use often, the software will not flag it as a possible error.

Another enormous benefit of being able to teach an ICR program “how to read” is that you can teach it to decipher data forms and hand-printed block letters.

For example, suppose your company sends out a customer survey card with its product. Perhaps this card has a set of multiple-choice questions, which the customer answers by marking off checkboxes. At the end of the survey, the customer is asked to fill in a series of boxes with the letters of his or her name and address.

An ICR system can be taught, as soon as the survey card is scanned, to record the answers to the questions, the customer’s name and address, and then ignore the rest of the form (to avoid storing redundant data). This information can then all be saved automatically in a database program.

Cost Advantages of OCR and ICR

OCR and ICR software can save you enormous amounts of time and money by minimizing the need for manual data entry and re-keying.

For example, an insurance agency could set up an ICR system to process claim forms. Data in specific fields on the forms could be extracted, related correspondence and forms could be generated automatically, and the information could all be archived in a fraction of the time it would take to read the forms, retype the data in the correspondence, and then file it away.

As another example, consider an office in which words are almost 100% of the data handled. Using an average of 300 words per page in a 60 page sample document, there would be 18,000 words. Manually typing 18,000 words would take a typist working at 50 words per minute about 6 *hours*. Scanning and performing OCR on that same document would take about 20 *minutes*. An OCR accuracy rate of only 95% would leave 900 words to be re-entered, or about 18 minutes worth.

A recent Yankee Group (Boston) study found that scanning systems can increase user productivity by 50%, and reduce data management costs by over 70%. Because of these benefits, installations of scanning systems in corporate networks is expected to increase by over 60% in the next decade.

Accuracy In OCR and ICR

Accuracy rates for OCR and ICR can vary anywhere from 50% accurate—which is awful—to 99% accurate—which is pretty darn good. The most important variable is the quality of the documents being scanned.

Obviously, if you try to scan a faded photocopy of a fax, you are not going to get good results. If, however, you scan a clean laser printout or a newspaper, your results should be excellent.

A 99% accuracy rate is adequate for most applications. However, some applications require *extremely* high accuracy. For example, if you are scanning a 5,000 page document, 1 word wrong out of every 100 means *many* incorrect words. In such situations, either close scrutiny of the scan process and the resulting bitmap image is necessary, or you must employ special, pre-programmed text validation techniques.

Errors in OCR and ICR can be caused by several conditions:

☞ Dirty or faded pages

☞ Pages with colored or patterned backgrounds

☞ Cursive text, and other fancy or “noisy” typefaces

☞ Skewed pages on the scanner or in a photocopy

☞ Touching characters

☞ Complex column layouts

☞ Scanning at resolutions less than 200 dpi

What Is Scanning Software?

Scanning software refers to software packages that assist in the scanning process by providing at least two sets of generic functions:

- ☞ Controls for operating your scanner
- ☞ Tools for defining active regions and graphical treatments to use on the pages to be scanned

Some scanning software also includes an OCR or ICR module (see above). In addition, some scanning software provides graphical image manipulation tools that can be used on images *after* they have been scanned. For example, tools that let you lighten, darken, or apply color effects to an image.

Depending on the type of scanner and job for which it used, scanning software may be some custom application that provides functions appropriate for a specific task. For example, the software used to drive the bar code scanners in a supermarket are highly specialized and have limited generic capabilities.

What Is Document Management Software?

Document management software is the controlling software that integrates:

- ✔ Scanning software
- ✔ Your favorite word processor, spreadsheet, and presentation programs
- ✔ Illustration packages
- ✔ Database software

In addition, document management software usually provides some sort of *workflow* control (see page 23) that lets you manage, index, print, archive, and distribute your scanned pages in meaningful ways.

What Are File Formats?

File format refers to the format in which data files are stored. In other words, when storing data as a series of bits, a set of rules is needed by the computer to understand the patterns in which those bits are organized.

For example, some file formats are designed to handle the binary representation of ASCII text; other formats are optimized to handle color photographs; still others are designed to handle information in a numerical spreadsheet.

Think of file formats as a type of *language* or *grammar*. If your program can only understand spreadsheet, then you must converse with it using files that are in spreadsheet format.

Types of Data File Formats

For desktop PCs, there are two general categories of data file formats, although there is some crossover between these two categories:

- ☞ *DOS-compatible* – Formats that work with IBM-compatible PCs running the MS-DOS operating system or one of its variants.
- ☞ *Macintosh-compatible* – Formats that work with Apple Macintosh-compatible computers running System software.

There are also UNIX file formats, but these are not discussed here because UNIX is used less frequently on desktop PCs than either DOS or Macintosh System.

Moreover, UNIX, being a generally more robust operating system, provides numerous data *filters* that allow data files from different operating systems to coexist on the same disk.

DOS-Compatible File Formats

When working with IBM-compatible (DOS) PCs, you can often figure out what format a given file is in by looking at the *extension* used in the file's name.

Specifically, DOS files are stored in what is commonly called the *8-dot-3* format, in which the first eight characters in a file name are just that—a name—and the last three (the extension) usually (but not always) indicate the file format; the name and the extension are separated by a dot (a period). For example:

```
filename.txt
```

The table on the next page lists common DOS file name extensions and their associated file formats. Please note that there are many other formats, and not all use specific extensions.

Extension	Format
.BMP	Bitmap format; the most common bitmap format used to represent graphical image files.
.DOC	Any of various document formats; used by word-processing and page layout programs such as Microsoft® Word™, FrameMaker™, and Lotus AmiPro™. Note that .DOC is not a generic format; different word-processing programs use their own .DOC format.
.DRW	Graphics file format used by drawing packages such as Microsoft Draw™ and Micrografx Designer™.
.DXF	Graphics format used by CAD/CAM programs such as AutoCAD®.
.EPS	Encapsulated PostScript®; a graphics file format understood by many graphics, page layout, and word-processing programs.
.FAX	Fax file format; like .DOC format, different fax programs use their own fax format.
.GIF	Graphics Interchange Format; a graphics file format used by online services such as CompuServe®.
.PCT	Macintosh® PICT format; a graphics format used by Macintosh computers.
.PCX	Another common bitmap format.
.RTF	Rich Text Format; a text format with embedded formatting codes.
.TGA	Truevision® Advanced Raster Graphics Adapter (TARGA); a high-resolution image and video format.
.TIF	Tagged Image File Format (TIFF); a bitmap format commonly used for scanned images. There are several varieties of TIFF files.
.TXT	ASCII text format.
.WKS	Lotus® 1-2-3®-compatible spreadsheet format.
.WMF	Microsoft Windows Metafile; another bitmap format.
.WPG	WordPerfect Graphics format; used by WordPerfect® word-processing software.
.XLS	Microsoft Excel®-compatible spreadsheet format.

Macintosh-Compatible File Formats

With most Macintosh-compatible file formats, there is no way to determine the format of a file just by looking at its name. Nor, unlike DOS, is there always a need to—Macintosh System software provides that kind of information to you in more reliable ways.

This is because Macintosh System software stores format information with each file, in an internal portion of the file called the *resource fork*. When you open a file, the application with which the file is compatible is automatically started—you don’t need to worry about where the file came from!

In addition, new Macintosh computers have *super drives*, which are diskette drives capable of reading files in DOS format. While these don’t solve all file format compatibility problems between the Macintosh and DOS environments, they are *extremely* convenient.

A Note About Binary Compatibility

When an application (a program, that is) is available on several operating systems, the file formats they use are often *binary-compatible*.

A binary-compatible format is one that can be used by a single application running under different operating systems. In short, if you can get a file on to a disk under a given operating system, then an application that supports binary compatibility across operating systems can use it—no conversion necessary.

For example, FrameMaker, from Frame Technology Corporation, is available on the UNIX, DOS/Windows, and Macintosh platforms. Files created in the Windows version are totally compatible with the UNIX and Mac versions.

What Is Indexing?

Indexing refers to the process of adding (keying in) search and retrieval information to files you create or scan. Similar to creating a library card catalog, indexing provides a data bank you can refer to later when trying to locate specific files or specific data in those files.

For example, when scanning a stack of 50 documents, you could add index information in the form of author name, keywords, description, date, time, and so forth. Later, when trying to locate one or more documents meeting certain search criteria, you can use you index entries to speed and narrow the search.

The specific index labels (the *fields*) you choose depend on the purpose for which the files are intended. For example, insurance claim forms could benefit from an index filed named `Claim Number`. A database of musical scores could benefit from a `Style` field (classical, rock, pop, country, and so forth).

Indexing tools are provided with document management software packages, like Lotus Notes®, Keyfile Corporation's Keyfile®, Emerald Gemsoft®, and the like. These packages also provide robust search and retrieval tools that take full advantage of the indexes you create.

A good indexing scheme is one of the most important components of an effective electronic document management system. As with paper indexes in a book, the more detailed your index scheme, and the more suited that detail is for the task at hand, the better your results will be.

Setting up a proper indexing scheme may take a little while to accomplish, but once it place, it will save you many hours and much money.

What Is Workflow?

In electronic document management, *workflow* refers to an automated organizational structure in which documents are electronically routed through a series of tasks or procedures.

For example, consider a bank office processing a client loan application. With an electronic document management system, the flow of events could be as follows:

❶ Scan the loan application.

❷ The document management software automatically marks and records the data from certain defined regions on the loan application, and then archives the application.

❸ The software is able to glean from fields on the application that this is an auto loan.

❹ The software notes that the dollar amount of loan is above a certain amount, and therefore initiates the series of sign-off procedures needed to approve the loan.

❺ The application is routed to the electronic mail In Box of the first loan officer on the list whose signature is needed.

❻ At the same time, electronic mail is sent to other loan officers on the list, notifying them that the loan is in process, and is awaiting signature from the first officer.

❼ From his or her computer, the first loan officer chooses the `Approved` button—other choices may be `Not Approved`, `Need More Info`, and so forth.

❽ The software sends the application on to the next officer, and updates other officers on the list.

❾ The signature process is tracked by the software, sometimes suspended pending officer approval, until the signature process is complete.

❿ Pending the results of the signature process, an approved or not approved letter is generated automatically and sent to the applicant.

Taking the Plunge

This section is where we get down to *practical* issues:

☞ Deciding how electronic document management can help you.

☞ How to implement an electronic document management system.

☞ Costs and benefits of electronic document management.

☞ Avoiding errors.

☞ Avoiding legal hassles.

The topics included in this section are listed below, along with the page numbers to which you can turn for specific information.

After reading this section, you may want to read the next section, “Success Stories,” starting on page 43, for examples of companies who have “taken the plunge” (and lived to tell about it).

The Problem

You’ve got problems. You’ve got at least one problem, anyway:

- ✔ You’re swimming in client records—billing information and payments, histories, documents, and exchanges—and you need to be able to find things among your records *quickly* and *accurately*.
- ✔ You have reams of paper, in cabinets and piles, around your offices. You need to put them *somewhere*, but you can’t get *rid* of them. You need to *remember* what’s in them. You need to find a clear space for your coffee mug on your desk.
- ✔ You are spending 85% of your time *processing* forms. You are spending the remaining 15% of your time *correcting* form processing errors.
- ✔ Your inventory keeps changing. Your product literature keeps changing, Stuff keeps getting moved from one location to another. You need to keep track of what’s where, when. You need to know what’s available *immediately*.
- ✔ You have microfilm. You have microfiche. Old, dusty bins filled with microfilm. You want to be able to use the images on this film in a more modern way—you want to view them, print them, *extract data* from them, sort through them efficiently, according to various search criteria.

If any of these problems sound familiar, then electronic document management will help you. Read on to find out how to get started.

Getting Started

The first step in implementing an electronic document management system is converting your old documents into an electronic format.

Before doing this, however, you need to decide what type of document conversion you want to undertake. Specifically, you may decide whether you want to:

- ☞ Scan and index only
- ☞ Perform optical character recognition (OCR) on the scanned documents
- ☞ Format your OCR output
- ☞ ICR your documents into database format

These four choices are explained below.

Scan and Index

The scanning process converts your documents into bitmap images (see page 8). Indexing happens during the scanning process, when index entries are manually attached (keyed in) to the scanned images (see page 22).

To create an effective index, you need to develop an indexing scheme that makes sense for your business. For example, an insurance group may want to index client names, policy types, phone numbers, account numbers, and so forth. A professional sports club may want to index athlete’s names, their salaries, the names of their lawyers, their bondsmen…

Simple scanning and indexing is best suited for general archiving. It allows rapid access to your documents, but does not retain any of the “intelligence” that may be contained in your documents. That is, you can retrieve a *picture* of a document, and then view it or print it, but you cannot edit or export the information in the document (text, for example) to a word-processing or database program.

Optical Character Recognition

If you choose this option, your scanned document images are processed through an optical character recognition (OCR) program (see page 13). This process converts the bitmap images into editable text—text that can be used in your word-processing or database programs.

This also makes it possible to perform *full-text* searches on your documents—that is, you can look for specific words in the document, whether they have been indexed or not.

☞ This level of OCR does not include formatting your scanned output (see “Formatted OCR,” below).

For example, if you scanned all documents relating to a particular court trial (*yes*, this has already been done with the O.J. Simpson case), and performed OCR on them, you could ask questions like: “Show me all plane reservations and motel charges for the defendant between 5pm and 11pm, from September 12, 1993 to June 15, 1994.” This is one reason why you see trial attorneys carrying portable PCs in court.

Formatted OCR

One of the strengths of OCR is that it lets you preserve much of a document’s original formatting—margins, columns, page layouts, text treatments (like font changes, bold, italics, underline), and so forth.

The process of *formatting* your OCR output makes your scanned documents compatible with your favorite word-processing and spreadsheet programs—a far more useful format than simple ASCII text.

In addition, you should be aware that OCR software does not produce 100% accurate results. Depending on the condition of the original documents, you can usually depend on an accuracy rate of 90% to 95%.

If this rate is not good enough, you must perform *manual cleanup* on the OCR'ed documents to correct any conversion errors.

The dual processes of formatting and performing manual cleanup can be labor-intensive, and require OCR operators with strong language skills for the language in which the documents are written, and solid knowledge of the format in which you want to save your documents.

OCR cleanup in particular, depending on the scanning initial results, can be relatively painless or quite a lot of work. In either case, it requires that the documents be proofread carefully. If you need 100% accuracy, however, you must allow time for this step.

ICR To Database

Intelligent Character Recognition (ICR) is similar to OCR, except that you can convert your scanned images into a format that is *directly* compatible with your favorite database program (see page 13).

For example, suppose your company sends out a customer survey card with its product. Perhaps this card has a set of multiple-choice questions, which the customer answers by marking off checkboxes. At the end of the survey, the customer is asked to fill in a series of boxes with the letters of his or her name and address.

An ICR system can be taught, as soon as the survey card is scanned, to record the answers to the questions, the customer's name and address, and then ignore the rest of the form (to avoid storing redundant data). This information can then all be saved automatically in a database program.

ICR puts the "intelligence" back into your scanned documents—though, of course, it won't *add* any intelligence that wasn't there in the first place…

Converting Your Old Stuff

After deciding the kind of conversion to electronic format you want to perform on your documents, the next step is actually start the conversion.

Even here, though, there are several questions you may want to ask:

- ☞ Should I convert all my old files?
- ☞ Should I do some or all of the conversion myself?
- ☞ How can I make the process cheaper and faster?

The answers to these questions are discussed below.

Should I Convert All My Old Files?

A common question is whether to convert all your old files (*backfiles*) in addition to your new ones (*day-forward*).

Most often, the best decision is consistent with the decision to move forward with an electronic document management system in the first place —that is, *yes, convert the whole lot.*

The point of electronic document management is to improve the diverse ways in which you do business—speed the handling of customer requests, answering (with the right answers, that is) complicated questions from customers, increasing worker productivity, and so forth.

Moving into the modern world of electronic document management while still relying on antique file drawers is usually a very awkward and expensive anachronism.

Should I Do Some or All of the Conversion Myself?

Your choice is to do it yourself, or to hire a scanning service bureau, like ScanCenters of America, to do it for you.

To do it yourself, you must be:

- ✔ Experienced in managing and organizing the workflow of large organizations
- ✔ Surrounded by a trained staff and management team
- ✔ Equipped with suitable scanning hardware and software

In addition, you will require workers for the conversion who:

- ✔ Have excellent language skills
- ✔ Can operate scanner and OCR equipment
- ✔ Know document management software
- ✔ Will work (perhaps part-time) for competitive wages.

Meeting all of these criteria is a daunting challenge for most personnel departments. Usually, managers such as yourself cannot find enough skilled workers to complete the job. The result is comprise: your workers perform the conversion as well as their own jobs—quality on both fronts suffers.

☞ It’s estimated that it costs twice as much, and takes twice as long, to do large-scale conversions yourself, rather than using a scanning service bureau.

How Can I Make the Process Cheaper and Faster?

There are a number of things you can do to facilitate the document conversion process:

- ✔ Spend extensive time thinking through all the details. Write down all of the rules you want to use during the conversion. For example, “How are Post-It notes handled?”
- ✔ If you use a consultant, have him or her talk to your scanning service bureau as early as possible in the process.
- ✔ Thoroughly understand the indexing scheme you want to use, and whether you’ll need OCR, formatted OCR, or ICR.
- ✔ Work with your staff and consultants on all the “what if” scenarios.
- ✔ Get rid of “junk” documents you don’t really need.
- ✔ Prepare your documents in-house before shipping them to your scanning service bureau. For example:
  - Remove all staples.
  - Mark each carton with a simple, effective serial number.
  - If you’ll need some of your documents while they are being converted, make copies for yourself.
  - If you are giving your scanning bureau documents in pieces (separate boxes), develop your own indexing scheme for the boxes so you’ll have an audit trail of where the boxes are (ScanCenters will maintain a “mirror” trail with you).

  ☞ Preparation time is usually 15% of the total conversion—preparation takes one to two hours per thousand pages.
- ✔ Do a trial run with your scanning service bureau to validate sample scans, file structure, and indexing schemes. Provide samples of all sizes and shapes of documents to be scanned.

Avoiding Errors

No technology is fool-proof. Scanners jam, read two pages at a time, and otherwise bedevil their operators. The results could be that certain pages of certain documents never make it into the computer's storage device—and could be lost *forever*…

Carefully select your scanning service bureau; make sure they use meticulous care and quality control to ensure that every page of every document you give them ends up in the right place on your storage device.

Costs and Benefits

More and more, sophisticated companies can justify the costs of converting to electronic document management on the basis of the answer to one question: "Will having these documents available on-line help us be more efficient and server our customers better?"

The answer to this question is almost always affirmative. In addition, once the conversion to electronic format is complete, the cost of using and maintaining an electronic document management system is *much* less than its paper counterpart.

Examples and explanations of these costs and benefits are provided on the following pages.

Merisel's Example

Merisel Corporation did an internal cost justification for their own electronic document management system in the Telemarketing and Order Entry departments. The table on the next page lists some of their findings.

Filing	Pages	Rate	Hours	@ Cost	$ Total
Manual	14,000	8 min.	1,866	13.00	24, 258
Electronic	14,000	5 min.	1166	13.00	15,158
				Savings Per Week	9,100
Retrieval	**Pages**	**Rate**	**Hours**	**@ Cost**	**$ Total**
Manual	5,000	15 min.	1,250	13.00	16,250
Electronic	5,000	1 min.	80	13.00	1,083
				Savings Per Week	15,167
	System Cost $	27,500		Total Savings/Week $	24,267

Merisel's system paid for itself in less than two weeks!

Some Typical Numbers and Costs

Consider the following typical costs in a manual (non-electronic) office environment:

- ✔ There are 3,000 pages in a typical four-drawer filing cabinet.
- ✔ Each filing cabinet occupies 9 square feet.
- ✔ Filing cabinets cost about $150 each.
- ✔ There are roughly $40 worth of file folders per cabinet.
- ✔ A bookshelf costs around $100.
- ✔ 18 minutes is the average manual search time for a document.
- ✔ $40 per hour is a professional rate and $7 to $11 for filing people
- ✔ 20% of all active documents are photocopied at $.05 per page.
- ✔ 6 minutes is the average time it takes to process a fax.

Other Cost Considerations

✔ The Gartner Group says that the cost of documentation and document maintenance is second only to payroll as the largest expense within most companies.

✔ Thanks to the PC and word-processing software, business is gradually moving into a world where documents are created electronically, delivered electronically, and printed only on demand. It is estimated that only 5% of all documents currently exist in electronic form.

✔ Consider the growth rate of CD-ROM drives:

- World-wide shipments of CD-ROM drives reached 17.45 million units in 1994, up 160% from 6.74 million, according to Dataquest of San Jose.
- Also, half of all CD-ROM titles were released in the last two years.
- Companies are buying phone lists, and stacks of anything that falls in the general category of reference materials.
- Of the Fortune 1000 companies, 91% plan to purchase CD-ROMs in the next 12 months.

Other Benefits of Electronic Document Management

Other benefits of an electronic document management system include:

- ✔ **Save Space –** The paper contents of an entire floor of paper document storage can be put into a space the size of a desk.
  - With a CD-ROM, you can put roughly 20,000 scanned document pages—or roughly 300,000 pages of ASCII text—on one drive. You can execute a search in about three seconds, and display a page in ten seconds. In another 30 seconds, you can print a copy.
  - A catalog listing the parts for every car model in the last 10 years would consume just seven CD-ROM disks—more than a million pages!

- ✔ **Search By Multiple Keys –** Try using the following criteria to manually search for documents: "Of all the automobile registrations we have, how many were issued between the 1st and 15th of November last year for people named Jones? With an electronic document management system, such a search would take about three seconds.

  Price, Waterhouse did a study to test the feasibility of an electronic document management system. They looked manually for 20 documents, selected randomly from a sample of 20,000. After 67 hours, they found 15. The computer found all 20 of them in 20 *milliseconds!*

✔ **Simultaneous Access –** Anyone on the system can access at any time any document for which they have access privileges.

✔ **File Continuity –** With a paper file, if someone removes a document and forgets to replace it—or puts it back in the wrong place—the next person will never find it. In an electronic system, once a document is filed, it stays filed.

✔ **Tracking Access and Use –** Using electronic files, management can keep track of how often people access certain records. If, for example, it is found that 20% of the records are being accessed 80% of the time, managers can optimize the storage of those files to further increase access speed and reduce cost.

✔ **Document Life –** Electronic storage on CD-ROM is good for 50 to 100 years. Paper gradually ages—or worse, gets smudged, torn, or catches fire.

✔ **Long Term Management –** Having all documents electronically stored makes policies about archiving and destroying documents easier to implement.

Disadvantages of Electronic Document Management

Yes, Virginia, there are (just a few) disadvantages to electronic document management. Some of these are:

✔ You become even more dependent on your computers. Then again, at least you can *carry* a notebook computer—try doing that with a filing cabinet!

✔ If a document gets misfiled (mis-indexed) when it is originally created in the system, it is going to be even more lost than in a paper filing system. *To err is human, but to really screw up takes a computer!*

✔ There will be some loss of image quality in the transformation from paper to scanned document. Depending on the techniques used in scanning and compression (and a lot of other variables), this may create a problem much later when recovering some stored documents. Most of the issues can be mitigated, however, by addressing them *before* the conversions are done.

Legal Considerations

Copyrights

In general, if you scan someone else's copyrighted material without their permission, you are breaking the law.

The copying by "exact means" of photographs, drawings, paintings, and other documents by photography, scanning, or photocopying can be an infringement of copyright laws. The person doing the unauthorized copying could be potentially subject to a number of remedies, including actual and statutory damages, impoundment of the materials, and other penalties.

☞ Get permissions from your sources first!

An exception to this is the concept of fair use—when the amount of the work being copied is not substantial, or is for a different use than the original (for example, not for profit).

Legal Documents

Scanned and electronically stored documents—especially on CD-ROM disks—are often admissible as court evidence. The law varies greatly from state to state, however, so check before assuming your files are acceptable.

☞ Electronic documents are accepted by the SEC, the IRS, and many other U.S. government agencies.

Be aware that there is a substantial, rapidly changing body of U.S. law, which is still coming to grips with the implications of electronic data, scanned documents, and intellectual property in general.

Some Notes About Microfilm

Microfilm technology is old, stable, and well-defined. If processed and stored in strict compliance with national standards, it has a very long life, and may be duplicated almost indefinitely, albeit with some generational loss.

It is also, however, unpopular due to is awkward, linear arrangement. Unlike computer disks, data on microfilm is stored sequentially; there is no *random access*. Also, it can usually be accessed by only one person at a time, does not convert easily to hard copy, and is not compatible with direct storage on a computer.

☞ Existing microfilm can be scanned and converted into digital images. ScanCenters of America offers such a service.

Computer Output to Microfilm/Microfiche

Computer Output to Microfilm/Microfiche (COM) is an older computer technology that stored images directly onto microfilm. COM was cheaper and more compact than paper. However, like traditional microfilm, the process created sequential files (rather than random access), which made on-line retrieval of the data slow and awkward. In addition, separate copies had to be kept in each user location, and only one person at a time could use them. COM files cannot be read by a computer.

Computer Output to Laser Disk

Computer Output to Laser Disk (COLD) is a common means of replacing microfilm (or existing magnetic tapes and other, older media) by converting the existing media to binary data, database-compatible formats, and text. These documents have “intelligence”: the data is often self -indexing, and can be easily stored and retrieved.

COLD and electronic document management can work together with other business forms and documents on your system. With COLD, you can scan a large amount of data in ASCII format, and store it with *one* forms definition. Later, you can create printouts of the form with any specified data from the database inserted—the forms thus look as they did when they were originally on paper.

There are several software packages for COLD, available from different vendors. To use one of these packages, you must check your system requirements to determine how COLD can fit in.

Another disadvantage to COLD technology is that there are no common computer-based standards, and each user must consequently make a decision regarding storage from a rapidly changing field of optical disks. The result is that each system ends up being “one of a kind.”

This section briefly describes how three companies “took the plunge” and implemented an electronic document management system.

Company A

A New England-based real estate developer sought a solution to a continuing paper problem they faced: for every property they owned, when it came time to refinance, mortgage, or enter into negotiations for sale, it was necessary to furnish to either the bank or prospective buyer an extensive list of approximately fifty critical items. This list included such things as loan documents, occupancy permits, environmental impact statements, as-built drawings, traffic reports, tenant leases, tenant build-out drawings, brokerage agreements, budgets and proformas, and insurance policies.

This information was typically maintained by a variety of different sources. Some information was kept in the developer's own files, but most of it resided at the Town Clerk's office, building contractors' offices, the bank, insurance companies, or in file boxes in off-site storage.

For one particular property closing, the developer budgeted $750,000 for legal fees alone. This was to cover the law firm's time to actually close the deal, but largely to gather and photocopy all of the requisite documents. There had to be a better way.

An electronic document imaging system was implemented to capture all of this information in one location. For one such project, there were some 600 drawings and 24,000 pages of text information. All were scanned and indexed into the imaging software in a period of weeks.

Now, not only does the firm have the information for critical transactions on the property, but also for day-to-day operations, which is saving them much time.

For example, all tenant leases were OCR'ed for full text retrieval. During the summer before the conversion to an electronic document system, a temporary legal aid was hired to review all lease documents and compile a list of all clauses relative to landlord or tenant termination, sorted by tenant. The temp worked for six weeks on the project. When the same task was performed on the electronic system, it was completed in 90 minutes—and the automated retrieval found four additional cases the temp had missed.

The next step in this developer's process is to archive project information to CD, so that the lending institution may have a complete copy of the property information. It is anticipated that this will not only make future transactions easier with the bank, as both parties have ready access to the information, but also bodes well for this developer's working relationship with the lender.

Company B

A large, multinational management consulting firm sought a better means of providing its fifteen field offices the tools to serve its clients. From a total of 13,000 presentations, generated by the firm over a twenty year period, 250 were chosen as the best representations of solutions the firm offered to its clients in a variety of vertical marketplaces and business conditions. These selected presentations would be used as models by all international offices when proposing strategies to clients.

The firm chose to scan the presentations, and then produce CD-ROMs with easy-to-use indexes that could be distributed and used by all employees. Each presentation was indexed by title, catalog number and study area. The user interface had to be accessible enough so that all employees could access the information and copy it, fax it, print it, or import it to Lotus Freelance for new presentations, either from individual workstations or across an network.

The 250 reports comprised approximately 14,000 pages of text, graphics and charts. When completed, they all fit on a single CD-ROM and employed a user-friendly retrieval and viewing interface.

As a result of this successful application, the firm is now looking at distributing, in a like fashion, all print articles written about the firm and by its employees as a marketing tool.

Company C

After implementing an electronic document management system, a Department of Public Works in a major U.S. city tripled its revenues from the collection of parking ticket fines.

The system processes and stores 8,000 to 10,000 tickets per day, from over thirty different issuing agencies. By the next day, these tickets can be displayed in 5 seconds or less on any of the DPW's 2,000 workstations.

Converting the existing backfile of 4.4 million tickets took about five months. Once converted, the data in these backfiles was available on all workstations throughout the system. The new system also contains correspondence from traffic offenders regarding tickets and other topics, which were originally submitted on all kinds of paper, and in numerous formats.

Glossary

Ablate: To remove. Used to describe the laser-readable "pits" in the recorded layer of optical disks.

Acetate-base film. A film substrate used in microfilm production. Considered a safety film (ANSI Standard).

Additive Color: All the colors in the light spectrum *add up* to make white light. Computer monitors use a three additive colors, Red, Green & Blue (RGB).

ADC: Analog to Digital converter. Changes analog signals to digital representations (numbers).

Aliasing: When computer graphics output has jagged edges or a stair stepped appearance when magnified. Homonym is "anti-aliasing".

AIIM: The **A**ssociation for **I**nformation and **I**mage **M**anagement – focused on electronic imaging.

Alphanumeric: Characters composed of letters, numbers (and sometimes punctuation marks). Excludes printer/flow control characters, (Carriage Return/XON & XOFF).

Analog: The electrical replica or waveform of a physical process caused by changes in amplitude or frequency. Opposite of digital (Zeros & Ones).

ANSI: American **N**ational **S**tandards **I**nstitute. Member of ISO and IEC.

Aperture Card: An IBM punch card with a window which holds a 35mm frame of microfilm. Indexing information is punched in the card.

ASCII: (pronounced *ask-ee*) **A**merican **S**tandards **C**ommittee **II**. An eight bit computer coding structure for letters, numbers and characters in which seven bits are used to identify each individual entity (128 maximum), with one bit for parity. When no parity bit is used, all eight bits can be used to represent up to 256 characters; this character set is *extended ASCII*.

Aspect Ratio: The relationship of the height and width of any image. This must always be preserved to prevent distortion.

AVI: Audio-**V**ideo **I**nterleave. A Microsoft standard for Windows animation files. The format interleaves audio and animation to provide medium quality multimedia.

Backfiles: Existing paper or microfilm files.

Bar Code: A method of representing data by combining lines of varying width (e.g.: UPC codes).

BBS: Bulletin **B**oard **S**ystem.

BPI: Bits Per Inch. For instance, this defines data densities in disk and magnetic tape systems.

BCS: Boston **C**omputer **S**ociety, one of the first associations of PC/Apple users and one of the largest and most active.

BIOS: Binary (or **B**asic) **I**nput **O**utput **S**pecification – the specific PC input/output "rules" and the programs which execute these to allow the transfer of information to/from the "central processing unit" of the PC.

BIT: Binary Dig**it**. Single position in base 2 arithmetic (2^n) – either on (1) or off (0).

Bit Map: Creating characters or images by creating a "picture" (matrix) of individual bits (pixels). The individual bits may just be binary (black and white) or high definition color. In color systems, the "z-axis" of each pixel has a value which represents the "shade of gray" or color of the bit. This value can be as high as 32 bits for very high resolution color. This results in a large, uncompressed file. For instance, a 300 dpi, E-Size drawing bit map is approximately 16MB.

BMP: Bit **M**a**p** unique format for Windows electronic graphics files.

Box: A square graphic element on a form used to enter a single character, usually used in strings for entering constrained data.

bps: bits **p**er **s**econd.

Buss: (*also* **Bus**) The "highway" which connects the various components of a computer system.

BYTE: Eight bits. The ASCII standard to define letters, numbers and characters – maximum of 256.

KB – Kilo-bytes, a thousand bytes (actually 2^{10} or 1024 bytes).

MB – Megabytes, a million bytes, (actually 2^{20} or 1,024 KB or 1,048,576 bytes)

GB – Gigabytes, a billion bytes (actually 2^{30} or 1024 MB or 1,073,741,824 bytes)

Cache: A dedicated, high speed portion of computer memory which can be used for the temporary storage of frequently used data to make the application run faster (prevents having to constantly access the data from disk/tape storage).

Captain Crunch: A compression algorithm, see MPEG.

CCITT: Consultative **C**ommittee for **I**nternational **T**elephone & **T**elegraphy. Sets standards for phones, faxes, modems etc. The standard exists primarily for fax documents.

CCITT Group 4: A compression technique/format that reduces a file generally, about 5:1 over RLE and 40:1 over bitmap. For example, at a 300 bpi scan rate, the approximate storage requirements are:

Size	Raw	RLE	Group 4
A	1MB	200K	40K
B	2MB	400K	75K
C	4MB	820K	150K
D	8MB	1.6MB	300K
E	16MB	3.2MB	580K

CCD: Charge **C**oupled **D**evice. A computer chip (with say 2048 cells) whose output is proportional to the light or color passed by it. Individual CCD's or arrays of these are used in scanners as a high-resolution, “digital camera" to "read" documents. These devices are micro-chip size and their resolutions run as high as 1000 pixels per inch.

CD: Compact **D**isk. A 4 3/4" diameter device which can be read by a laser beam.

CDMA: Code-**D**ivision **M**ultiple **A**ccess – an emerging wireless communication technology for all digital voice and data networks.

CDPD: Cellular **D**igital **P**acket **D**ata. A data communication standard which uses the unused capacity (bandwidth) of cellular voice providers.

CD-R: Compact **D**isk **R**ecordable. The standards for recording CD-ROM disks. The digital disks are 4" in diameter and can store 650 MB.

For standard CD's. Each disk has a layer of laser sensitive, dyed polymer plastic sandwiched against a reflective layer between protective layers. When the laser burns a spot in the polymer, the reflective surface shows through the hole.

Dye polymer is easier to burn and requires a much lower power laser than to burn holes through a metal layer (such as ablative optical WORM drives.) The CD-R media is gold in appearance, rather than silver surface of a typical CD-ROM.

The *logical format* standard is ISO (International Standards Organization) 9660. There are several standard formats:

1. "Yellow Book" – for simple computer data or images. Divides the tracks into 2,352 byte-sectors of which 2,048 hold data and 304 bytes are devoted to headers, mode selection and error correction.
2. "CD-ROM-XA" (Extended Architecture) or "Mode 2" for interleaving data, audio and video on the same disks. Mode 2 sacrifices error correction for a larger usable data storage. Sectors have 2,336 bytes of data space.
3. "Red Book" or "CD-Digital Audio" for digitally sampled audio, technically PCM 44.1 kHz, sampled 16-bit stereo audio. The standard for recording music.
4. "Orange Book" or "Multisession". The standard that software follows to encode a blank CD. Part I is the standard for "rewritable" (MO) CD-ROMs . Groups of data can be added to the disk at different times.
5. "Green Book" or "CD-1". For interactive games or video.

CD-Recordable: Often also used as an acronym for CD-ROM's that can be written more than once. The succeeding writings must utilize unused sections of the original, with a library o directory of the total use. Optical storage technology using formats compatible with CD-ROM's. CD-ROM discs must be "pre-mastered" to insure that the data is correctly formatted. Using a "double speed" recorder, it takes about a half hour to burn a complete 650MB disc.

CD-ROM: Compact Disk – Read Only Memory. A type of high density optical disk with a 4" diameter and a 650MB capacity. The information (1's or 0's) is permanently etched by a laser into the surface of the disk and read by a laser beam. The ISO 9660 standard defines how a CD-ROM is written for computer interface. It is not rewritable. It is legally accepted and written on a single-side.

Centronics Interface: A parallel interface standard for connecting printers and other devices to computers. Pioneered by the Centronics Inc., a printer manufacturer in New Hampshire. Uses a 36 pin connector. See SPP.

CGA: Color **G**raphics **A**dapter. (See VGA).

CIE: Commission **I**nternational de l'**E**clairage. The international commission on color matching and illumination systems.

Cine-Mode: Data recorded on a film strip such that it can be read by a human when held vertically.

Cinepak: A compression algorithm, see MPEG.

CITIS: Contractor **I**ntegrated **T**echnical **I**nformation **S**ervice. The Department Of Defense now requires contractors to have an electronic document image *and* management system

Client/Server: A computer system functionally distributed across several nodes on a network, sometimes called a distributed application. The basic theory is that the various components of the system can be tailored to perform specific functions, hopefully for the good of the entire network. Client/Server systems are also typified by a high degree of parallel processing across distributed nodes. Usually the clients are individual PC's connected to server(s) which act as central storehouses and "traffic cops" for information and applications.

COLD: Computer **O**utput to **L**aser **D**isk. The computer system contains files of ASCII data (from input or application programs) or bit-mapped files previously scanned from microfilm documents or pictures. These output files are compressed by a factor of 5-20:1 from the original documents and stored on WORM optical/laser disks. The stored data is then available to all on the network. Generally, the format of these databases are compatible with SQL and imaging formats.

COM: Computer **O**utput to **M**icrofilm. The computer converts and stores data directly on microfilm/fiche from a variety of available inputs. This older technology is cheaper and more convenient than paper, but one of the most difficult to use in actually storing and retrieving the data.

Comb: A series of boxes with their top missing. Tick marks guide text entry. Used in forms processing rather than boxes.

Comic Mode: Human-readable data, recorded on a strip of film which can be read when the film is moved horizontally to the reader.

Composite Video: A video stream that combines red, green, blue and synchronization signals into one so it only requires one connector. Composite video is used by most televisions and VCR's.

Component Video: Separate luminosity and color signals that provide the highest possible signal quality. Distinct from video standards such as NTSC or PAL.

Compression: Any method which reduces the amount of data necessary to transmit information from one point to another. Compression generally eliminates redundant information and/or predicts where changes will occur. "Lossless" compression techniques totally preserve the integrity of the input. "Lossy" methods disregard some of the originals.

CMYK: Cyan, **M**agenta, **Y**ellow and Blac**k**. A subtractive method used in four color printing and Desktop Publishing.

Continuous Tone: An image (e.g.: a photograph) which has all the values of gray from white to black.

Convergence: Where the RGB signals "converge" on a single pixel. That pixel should be white at full brightness of the RGB components.

CPI: Characters **P**er **I**nch.

CPU: Central **P**rocessing **U**nit – The portion of a computer which performs most of the logical and arithmetic functions.

CRC: Cyclical **R**edundancy **C**hecking. Used in data communications to create a checksum character (hexadecimal) at the end of a data block.

CYAN: A colored ink. Reflects blue & green & absorbs red.

DAC: Digital to **A**nalog **C**onverter. Changes digital numbers to an electrical waveform.

DAD: Digital **A**udio **D**isk – "compact disk".

DAT: Digital **A**udio **T**ape – Although generally used for audio, a DAT (120 meters long) can hold up to 10 gigabytes if used for digital data storage. Has the disadvantage of being a serial, rather than a random access device.

DB: Data **B**ase. Information arranged in the computer in a rigorous, defined format to allow ease of recording and retrieval.

Descenders: the portion of a character which falls below the main part of the letter (e.g.: g, p,q)

DIA/DCA: Document **I**nterchange **A**rchitecture. An IBM standard for transmission and storage of voice, text or video over networks.

Digital: A system of mathematics consisting solely of zeros and ones. The mathematics used by digital computers. Used to represent characters and numbers and to mathematically manipulate these.

Digitize: The process of converting an analog value into a digital (numeric) representation.

Disk/Disc: Round, flat storage media with layers of material which enable the recording of data.

Disc: An optical disc.

Disk: A magnetic floppy or hard disk.

Dithering: Manipulating the arrangement or shape of dots to simulated gray tones. (e.g.: Newspaper pictures).

DOCUMENT SIZES (U.S.):

A Size	8.5" by 11"	(A4)
B Size	11" by 17"	(A3)
C Size	17" by 22"	(A2)
D Size	24" by 36"	(A1)
E Size	36" by 48"	(A0)

Dot Pitch: Distance of one pixel in a CRT to the next pixel on the vertical plane. The smaller the number, the higher quality display.

DPI: Dots **p**er **i**nch.

DRAM: Dynamic **R**andom **A**ccess **M**emory, a memory technology which is periodically "refreshed" or updated – as opposed to "static" RAM chips which do not require refreshing. The term is often used to refer to the memory chips themselves. Varieties are:

CDRAM	Cache DRAM (contains static cache)
EDODRAM	Extended data out DRAM
EDRAM	Enhanced DRAM (contains a static memory buffer and cache controller)
SDRAM	Synchronous DRAM (added clock and burst addressing capability)
SGRAM	Synchronous Graphics RAM (a single port SDRAM)
WRAM	Window RAM (dual port video RAM)
VRAM	Video RAM (a dual ported DRAM, good for graphics)

DSP: Digital **S**ignal **P**rocessor (Processing) – a special purpose computer (or technique) which digitally processes signals and electrical/analog waveforms.

DTP: Desk**t**op **P**ublishing. PC systems used to prepare direct print output or output suitable for printing presses.

EDI: Electronic **D**ata **I**nterchange. Eliminating forms altogether by encoding the data as close as possible to the point of the transaction. (e.g.: Paying your phone bill direct from your PC to the system used by the phone company.)

EDMS: Electronic **D**ocument **M**anagement **S**ystems.

EISA: Extended **I**ndustry **S**tandard **A**rchitecture. One of the standard busses used for PC's.

EGA: Extended **G**raphics **A**dapter. See VGA.

EIA: Electronic **I**ndustries **A**ssociation – a trade association.

EIM: Electronic **I**mage **M**anagement.

Electrostatic Printing: Paper is exposed to electron charge. Toner sticks to the charged pixels.

Em: In any print font or size is equal to the width of the letter "M" in that font and size.

En: Half the width of an Em.

Endorser: A little printer in a scanner that adds a document-control number to each scanned sheet. Some forms control processing software can control this printer.

Encryption: The coding of messages to increase security and make transmission only readable by recipients with the ability to decode only by using the same algorithms.

EOF: End **o**f **F**ile. A distinctive code which uniquely marks the end of a data file.

EPP: Enhanced **P**arallel **P**ort – also known as Fast Mode Parallel Port. A new, industry standard parallel port, having high transfer times competitive with SCSI.

EPS: Encapsulated **P**ost**S**cript. Uncompressed files for images, text and objects. Only print on PostScript printers.

ESDI: Enhanced **S**mall **D**evice **I**nterface. A defined, common electronic interface for transferring data between computers and peripherals, particularly disk drives.

FAT: File **A**llocation **T**able – An internal data table on DOS-based disks that lists the contents and address of each file on the disk.

FAX: Short for facsimile. A process of transmitting documents by scanning them to digital, converting to analog, transmitting over phone lines and reversing the process at the other end and printing. "Group 3" indicates the 3rd generation of faxes which transmits a page at 9600 baud in about a minute – with a normal resolution of 203 x 98 dpi and a fine resolution of 203 x 196.

Forms Routing: The process of routing a form throughout an organization *electronically* – without any paper copies.

Fiber Optics: Transmitting with light pulses over cables made from thin strands of glass.

Field Separator: A code, usually a comma, that separates the fields in a record. (Also, a "delimiter")

FTP: File **T**ransfer Protocol. An Internet protocol to move files from one computer to another.

Full Duplex: Data communications devices which allow full speed transmission in both directions at the same time.

Full Text Search: The ability to search a data file for specified key(s) defined by the occurrence of words, numbers and/or combinations or patterns thereof.

GIF: A compressed file format used by the CompuServe system for photographs. Limited to 256 colors.

Gigabyte: A billion bytes or 1,000 megabytes (See "BYTE").

Gray Scale: The binary range of a graphic representation between pure black and pure white. A scale of 256 shades of gray will be a better representation than 16 shades.

Groupware: Software designed to operate on a network and allow several people to work together on the same documents and files.

GUI: Graphical **U**ser **I**nterface, or "gooey". Presenting an interface to the computer user comprised of pictures and icons, rather than words and numbers.

Half Duplex: Transmission systems which can send and receive, but not at the same time.

Halftone: The graphic representation of an object by dots, which simulate continuous tones. Usually used to represent or replicate an original photograph input.

HD: High **D**ensity (Floppy Disks) – A 5.25" holds 1.2 MB and a 3.5" holds 1.4 MB.

Hexadecimal: A number system with a base of 16 (2^4), 4 bits. The position digits are 0-9, A-F, where F equals the decimal value, 15.

Host: In a network, the central computer which controls the remote computers and holds the central databases.

HP-PCL & HPGL: Hewlett-Packard graphics file formats.

Hub: A central unit that repeats and/or amplifies data signals being sent across a network.

HTML: A **H**yper**t**ext **M**arkup **L**anguage, developed by CERN of Geneva, Switzerland. The document standard of choice of Internet. (HTML+ adds support for multi-media.)

Icon: In a GUI, a picture or drawing which is activated by "clicking" a mouse to command the computer program to perform a predefined series of events.

ICR: Intelligent **C**haracter **R**ecognition. The conversion of scanned images (bar codes or patterns of bits) to computer recognizable codes (ASCII characters and files) by means of software/programs which define the rules of and algorithms for conversion.

IDE: Integrated **D**rive **E**lectronics – An engineering standard for interfacing PC's and hard disks.

IEEE: Institute of **E**lectrical and **E**lectronic **E**ngineers. An international association which sponsors meetings, publishes a number of journals and establishes standards.

Image Processing: To capture an image or representation, enter in a computer and process and manipulate it.

Index: Creating a set of rules and data files which define scanned document sets and allow easy and complete retrieval.

Interlaced: TV & CRT pictures must constantly be "refreshed". Interlace is to refresh *every other* line once/refresh cycle. Since only half the information displayed is updated each cycle, interlaced displays are less expensive than "non-interlaced". However, interlaced displays are subject to jitters. The human eye/brain can usually detect displayed images which are completely refreshed at less than 30 times per second.

Internet: A worldwide computer network containing a broad array of services and information available to any individual with a PC and the paid connection.

ISA: Industry **S**tandard **A**rchitecture.

ISDN: Integrated **S**ervices **D**igital **N**etwork. An *all digital* network which can carry data, video and voice.

ISO: International **S**tandards **O**rganization.

JMS: Jukebox **M**anagement **S**oftware.

JPEG: A compression algorithm for still images, see MPEG.

Juke-Box: Automated disk changer for high-performance, centralized storage for multifunction CD-ROM's & optical disks

K: Generally accepted as shorthand for 1,000. Actually stands for 2^{10} or 1,024.

Kerning: Adjusting the spacing between two letters from the "normal" spacing. Often done to enhance the quality of the typography – for instance in a headline.

Kofax Board: The generic term for a series of image processing boards manufactured by Kofax Imaging Processing. These are used between the scanner and the computer, and perform realtime image compression and decompression for faster image viewing, image enhancement, and corrections to the input to account for conditions such as document misalignment, "speckles," etc.

LAN: Local **A**rea **N**etwork – usually a collection of PC's, connected by cable.

Landscape Mode: The image is represented on the page or monitor such that the width is greater than the height.

Laser Disk: Same as an optical CD, except 12" in diameter.

Latency: The time it takes to read a disk (or jukebox), including the time to physically position the media under the read/write head, seek the correct address and transfer it.

Leading/"Ledding": The amount of space between lines of printed text.

Line Screen: The number of half-tone dots that can be printed per inch. As a general rule, newspapers print at 65 to 85 lpi, large city newspapers at 100 or 120 lpi; magazines at 133 or 150 lpi; and, glossy, "coffee table" books at 175 to 200.

LZW: Lempel-**Z**if & **W**elch. A common, lossless compression standard for computer graphics – used for the majority of TIFF files. Typical compression ratios are 4/1.

MCA: Micro **C**hannel **A**rchitecture – an IBM buss standard.

Magenta: Used in four color printing. Reflects blue & red and absorbs green.

Mastering: Making many copies of a CD-ROM from a single master.

MDE: Magnetic **D**isk **E**mulation. Software that makes a jukebox look and operate like a hard-drive such that it will respond to all the I/O commands ordinarily sent to a hard drive.

Megabyte: A million bytes. See "byte".

MICR: Magnetic **I**nk **C**haracter **R**ecognition. The process used by banks to encode checks.

Microfilm: Film on which documents etc. are photographically greatly reduced in size.

Microfiche: Reduced sized document(s) filed on sheet microfilm (4" by 6"), containing reduced images of 270 pages or more in a grid pattern. Usually with a human-readable title.

MO: Magneto-**O**ptical. A disk storage technology which competes with traditional magnetic hard disks. Form factors are 3.5", 5.25" and 12". Advantages are that one 5.25" MO drive can store about 1.3GB (3 1/2" hold up to 230MB); media is removable and portable; and, can last for 20 years – ideal for archival storage. The disadvantages are cost, traditionally slower disk access and longer disk write times. The information is written on the disk by changing the polarity with strong magnets and read by a laser by sensing the magnetic flux changes (1's or 0's). This technology is re-usable.

MODEM: Modulator/**Dem**odulator. A device which can take digital data from a computer, translate it into analog signals (tones) and transmit the information over telephones lines. Another modem at the receiving computer will receive the information, translate it back from analog to digital and store it. Typical speeds are from 1,200 to 14,400 bits per second. Some modems also correct any errors which occur in the transmission process.

Monochrome: Displays capable of only two colors, usually black & white.

Mosaic: A program used for finding and reading documents on the World-Wide-Web.

MPEG-1 & 2: Two different standards for full motion video to digital compression/decompression techniques advanced by the **M**oving **P**ictures **E**xperts **G**roup. MPEG-1 compresses the bandwidth needed for 30 frames/second of full-motion video (several hundred megabytes) down to about 1.5 Mbits/sec. MPEG 2 only compresses to about 3 Mbits and provides for better image quality when comparing compressed files of the same size. This industry application competes with other compression techniques, know as JPEG, Captain Crunch, Cinepak and Indeo.

MS-DOS: Microsoft (**MS**)-**D**isk **O**perating **S**ystem. Used in PC's as the control system.

MTBF: Mean **T**ime **B**etween **F**ailure. Average time between failures. Used to compute the reliability of devices/equipment.

MTTR: Mean **T**ime **T**o **R**epair. Average time to repair. The higher the number, the most costly and difficult to fix.

Multisynch: Analog video monitors which can receive a wide range of display resolutions, usually including TV (NTSC). Color analog monitors accept separate red, green & blue (RGB) signals.

Non-Interlace: When each line of the video image is scanned separately. Computer monitors use non-interlaced video.

NTSC: **N**ational **T**elevision **S**ystem **C**ommittee. The North American TV standard – analog, 525 lines @ 30 frames per second. TV's line scan rate is then 15,750 lines per second (525 lines @ 30 Hz).

OCR: Optical **C**haracter **R**ecognition. The computer conversion of scanned input images (bar codes or patterns of bits) to computer recognizable codes (ASCII letters, numbers and characters).

OEM: Original **E**quipment **M**anufacturer – Classically, a company who buys products from another company, re-labels the products under its own name and re-sells (usually in large quantities). Has come to define nearly any large customer who re-sells products, branded or not.

OLE: Object **L**inking and **E**mbedding. A feature in Microsoft's Windows which allows each section of a compound document to call up its own editing tools or special display features. This allows for combining diverse elements in compound documents.

PAL: Phased **A**lternative **L**ine, the TV standard used in most of European. PAL uses 625 lines per frame and 25 frames per second – versus 30 for NTSC, resulting in more flicker.

PackBits: A compression scheme which originated with the Macintosh. Suitable only for black & white.

Packet: A fixed block of data transmission which also contains identity and routing information.

Paper Styles & Definitions:

a. Acid Free Paper – Won't change color (yellow) for many years.

b. Brightness – The percentage of light the paper reflects. Most white papers reflect 60% to 90%.

c. Coated Papers – "glossy" paper, coated with clay.

d. Cotton "Rag" Paper – Premium paper with 25% to 100% cotton fibers.

e. Laid finish – Paper surface embossed with lines to resemble handmade paper.

f. Ream – 500 sheets.

g. Vellum finish – A less smooth version of real vellum (fine parchment).

h. Wove finish – Very smooth surface. Characteristic of the majority of papers made.

Parallel: Transmission of all the bits (e.g.: in a character) at the same time. If the character has eight bits, there are eight wires. Faster and more expensive than serial where the eight bits would be sent, "sideways", one at a time.

PCI: Peripheral **C**omponent **I**nterface (Interconnect). A high-speed interconnect local bus used to support multimedia devices. Promoted by Digital among others.

PCMCIA: Personal **C**omputer **M**emory **C**ard **I**nternational **A**ssociation. Plug-in cards for computers (usually portables) which extend the storage and/or functionality.

PCX: The file format used for drawings by CorelPaint and Windows Paint Brush.

PDA: Personal **D**igital **A**ssistant – a small, usually hand-held, computer which "assists" business tasks.

PICA: One sixth (1/6) of an inch. Used to measure graphics/fonts.

PICT: Picture Format. A color file format exclusively for Macintosh.

Pitch: Characters (or dots) per inch, measured horizontally.

PIXEL: Picture **El**ement. One step/addressable position in the total TV or CRT presentation. The minimum VGA display has 307,200 pixels (640 by 480).

PMS: Pantone **M**atching **S**ystem. A color standard in printing.

POD: Print On Demand. Document images are stored in electronic format and are available to be quickly printed and in the exact quantity required, long or short runs.

Portable Document Format: A file standard for documents that can be processed (generally viewed and printed) by any computer, regardless of the specific application program which created the original.

Portrait Mode: A display where the height exceeds the width.

Proximity Search: For "full-text" searches, the ability to look for words which are within a prescribed distance of another word (e.g.: Find "glove" within 15 words of "baseball".)

QBIC: Query **B**y **I**mage **C**ontent. An IBM search system for stored images which allows the user to sketch an image and then search the images files to find those which most closely match. The user can specify color and texture – such as sandy beaches or clouds.

QIC: Quarter **I**nch **C**artridge. Digital recording tape, 2000 feet long, with an uncompressed capacity of 5 GB.

RAID: Redundant **A**rray of **I**nexpensive **D**isks. Arrays or Jukeboxes of CD-ROM's or CD-R's. There are five commonly uses, different *levels* of data protection, RAID 1 through RAID 5 which are tradeoffs of protection versus storage capacity.

Level 0: Data written in blocks across multiple drives without an protection on failures.

Level 1: Disk Mirroring.

Level 3: The drive spindles are synchronized such that the heads all seek at the same time and are positioned over the same read/write areas simultaneously. Data is written one bit at a time with parity to a separate drive. Thus if there were four disks in the array and there was a megabyte of data to transferred at 1 MB/sec, the effective rate is 4MB/sec.

Level 5: Writes data in chunks (usually smaller blocks 512 bytes to 2 K) with the parity striped along with the data. Achieves a higher I/O rate.

RAM: Random **A**ccess **M**emory – Memory which can be read or written in *any* section with one instruction sequence. (See DRAM)

Raster Display/Graphics: Represents images by an horizontal and vertical array of dots or pixels.

Recycled Paper: Federal Guidelines suggest at least 50% "non-virgin" content.

Refresh Rate: How many times a second and image on a CRT or TV is updated.

Registration: Lining up a forms image to determine which fields are where. Also, entering pages into a scanner such that they are correctly read.

RGB: Red, **G**reen and **B**lue. The three primary colors in the additive color family which create all the computer color video signals for a computer's color terminal.

Rewriteable Technology: Storage devices where the data may be written more than once – typically hard drives, floppies and optical disks. The assets are re-use, high speed and capacity. The optical disks have the same basic characteristics as a CD-ROM, except that you can write over the existing data.

ROM: Read **O**nly **M**emory – random memory which can be read but not written (i.e.: changed).

Rotary Camera: In microfilming, the papers are read "on the fly" with a camera that's synchronized to the motion.

Sampling Rate: The frequency at which analog signals are converted to digital values during digitization. The higher the rate, the more accurate the process. In printing: The number of pixels scanned per half tone dot.

SCSI: Small **C**omputer **S**ystem **I**nterface. A common, industry standard, electronic interface (highway) between computers and peripherals, such as hard disks, CD-ROM drives and scanners.

Serif: The little cross bars or curls at the end of strokes on type fonts. For example, in this sentence, the horizontal line at the bottom of the letter ‘r’.

SGML: Standard **G**eneralized **M**arkup **L**anguage. An *informal* industry standard (*lingua franca*) for open systems document management which specifies the data encoding of a document's format and content.

SGML/HyTime: A multimedia extension to SGML, sponsored by DOD.

SIMM: **S**ingle, **I**n-Line **M**emory **M**odule – A mechanical package (with "legs") used to attach memory chips to printed circuit boards.

SLIP: Serial **L**ine **I**nternet **P**rotocol. A connection to Internet in which the interface software runs in the local computer, rather than Internet's.

Smart Card: A credit card size device which contains a microprocessor, memory and a battery.

Splatter: Data that should be kept on one disc of a jukebox goes instead to multiple platters.

SPP: Standard Parallel Port (IBM-Centronics). See Centronics.

SQL: Structured **Q**uery **L**anguage, a standard 4GL programming language.

Subtractive colors: Since the colors of objects are white light *minus* the color absorbed by the object, they are called subtractive. This is how ink on paper works. The subtractive colors of process ink are CMYK (Cyan, Magenta, Yellow and Black) and are specifically balanced to match additive colors (RGB).

SVGA: "**S**uper" **V**ideo **G**raphic **A**dapter – one which exceeds the minimum VGA standard of 640 by 480 by 16 colors. Can reach 1600 by 1280 and 256 colors.

Telephony: Converting sounds into electronic signals for transmission.

Terabyte: A trillion bytes, or more correctly 1,024 megabytes.

TIF/TIFF: Tagged **I**mage **F**ile **F**ormat. The "de facto" electronic/computer standard for scanned, bit-mapped images – 8 bit color and gray scale. Originated in 1986 as a joint project of Microsoft and Aldus. Includes several types and groups which are compressed & uncompressed.

TGA: Tar**ga** format. This is a "scanned format" – widely used for color-scanned materials (24-bit) as well as by various "paint" and desktop publishing packages.

True Resolution: The "true" optical resolution of a scanner is the number of pixels per inch (without any software enhancements).

TWAIN: Tool Kit **W**ithout **A**n **I**nteresting **N**ame – a universal toolkit with standard hardware/software drivers for multi-media peripheral devices.

Typeface: There are over 10,000 typefaces available for computers. The general categories are:

Oldstyle: Faces have slanted serifs, gradual thick to thin strokes and a slanted stress (the "O" appears slanted)

Modern: Faces have thin, horizontal serifs, radical thick to thin strokes and a vertical street (the "O" does not appear to slant.)

Slab Serif: Faces have thick, horizontal serifs, little or no thick-to-thin in the strokes and a vertical stress (the "O" appears vertical).

Sans Serif: Faces have no serifs.

Script: From elaborate handwriting styles to casual, freeform, unconnected letter forms.

Decorative: Unusual fonts, designed to be very different and attention getting.

Ultrafiche: Microfiche which can hold 1,000 documents/sheet as opposed to the normal 270.

UNIX: A software operating system. Originally pioneered by Bell Labs – now widely used by workstations.

V.32bis: The ITU standard for 14.4 kbs modem communications.

V.34: The proposed ITU standard for 28.8 kbs modem communications.

VAR/VAD/VASD: **V**alue-Added **R**eseller/**V**alue-**A**dded **D**ealer/**V**alue-**A**dded **S**pecialty **D**istributor. Companies or people who sell computer hardware or software *and* "add-value" in the process. Most usually the value added is specific technical or marketing knowledge and/or experience.

Vector: Representation of graphic images by mathematical formulas. For instance, a circle is defined by a specific position and radius.

VESA: **V**ideo **E**lectronics **S**tandards **A**ssociation – concentrates on computer video standards.

VDT: Video Display Terminal – generic name for all display terminals.

VGA: **V**ideo **G**raphics **A**dapter. A PC industry standard, first introduced by IBM in 1987, for color video displays. The *minimum* dot (pixel) display is 640 by 480 by 16 colors. Then "Super VGA" was introduced at 800 x 600 x 16, then 256 colors. VGA can extend to 1024 by 768 by 256 colors. Replaces EGA, an earlier standard and the even older CGA. Newer standard displays can range up to 1600 by 1280.

WAN: Wide Area Network. Generally a network of PC's, remote to each other, connected by telecommunications lines.

.WAV: File extension name for Windows sound files. Compression is not required. .WAV files can reach 5 Mbytes for one minute of audio.

Workgroup: A group of computer users connected to share individual talents and resources as well as computer hardware and software – often to accomplish a team goal.

WORM: **W**rite-**O**nce, **R**ead-**M**any – Data storage devices (e.g.: CD-ROM's) where the space on the disks can *only* be written *once*. The data is *permanently* stored. This is often today's primary media for archival information. Disk sizes run from 5.25" (1.3 gigabytes) to 12" (8 to 10 gigabytes) capacities. There is also a 14" disc (13 to 15 gigabytes), only manufactured by Kodak's optical storage group. WORM's can also be configured into jukeboxes. There are various technologies:

Technology	Description	Benefit	Drawback
Ablative	Laser burns holes in disk	Unalterable data	Dust, moisture may affect media
Bubble-forming	Laser forms bubbles in the media	Unalterable data	Few drives available
Dye Polymer	Laser heats dyed layer to form bumps	Potential low media cost	Laser mechanism more expensive; disks wear out faster; few drives available
Magneto-	Laser focuses magnetic field	Many suppliers, long disk life	No true WORM in multi-function; data theoretically alterable
Phase change	Laser heat changes disk's molecular structure	One-pass data (no erase step)	Same as Dye Polymer

From Imaging Magazine, September, 1994

The expected viable lifetime of a WORM is at least 50 years. Since it's impossible to change, the government treats it just like paper or microfilm and it is accepted in litigation and other record-keeping application.

On the negative side, there is no current standard for how WORM's are written. The only ISO standard is for the 14" version, manufactured only by one vendor. A 5.25" standard is emerging from the European Computer Manufacturing Association but is not yet accepted. Further, WORM discs are written on both sides, but there are currently no drives that read both sides at the same time.

As for speed, WORM is faster than tape or CD-ROM, but slower than magnetic. Typical disk access times run between 40 and 150 milliseconds (compared with 11 ms for fast magnetic disks and 300 ms for CD-ROM. Data transfer rates run between 1 and 2 MB/sec (compared with 5 to 10 for magnetic discs and 600KB/sec for CD-ROM.

WYSIWYG: "**W**hat **Y**ou **S**ee **I**s **W**hat **Y**ou **G**et" – Display & software technology which shows on the computer screen exactly what you'll get when you print that screen. Usually requires a large, high-density monitor.

X.25: A standard protocol for data communications.

"XEROX" Printing: A beam of light hits an electrically charged drum and causes a discharge at that point. Toner is then applied which sticks to the non-charged areas. Paper is pressed against the drum to form the image and is then heated to dry the toner. Used in laser printers and copying machines.

Index