Document Imaging & Workflow Glossary
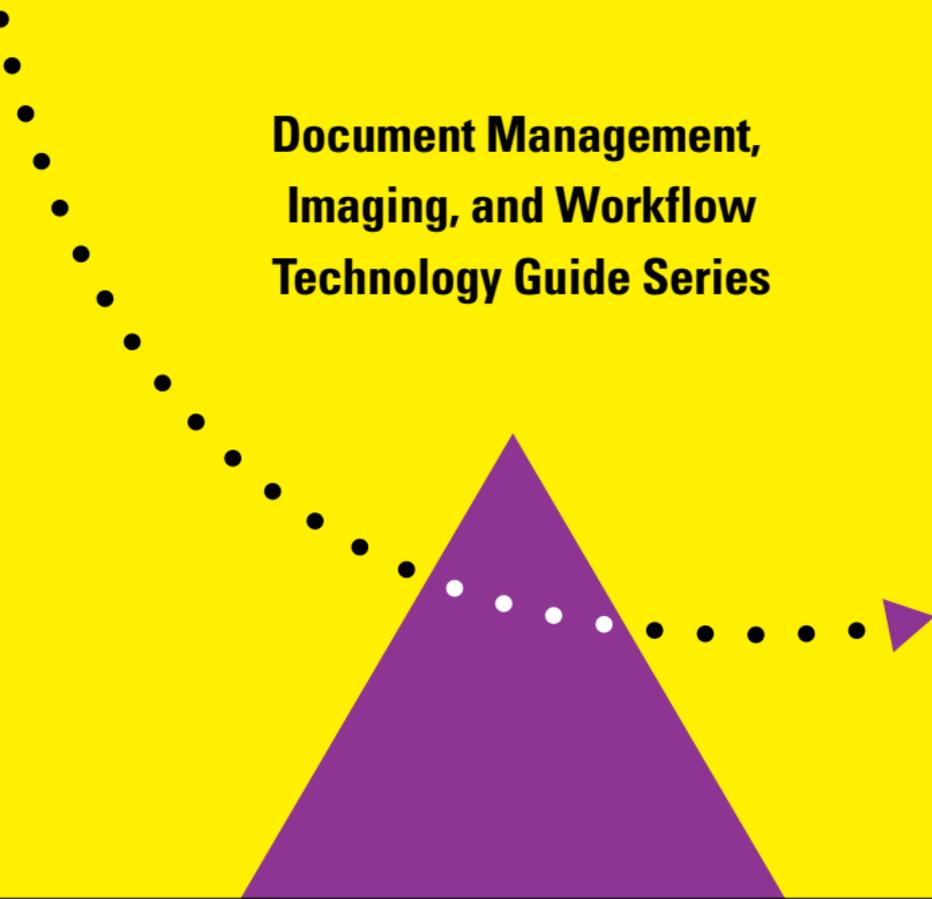
Document Management, Imaging, and Workflow Technology Guide Series

Abstraction—Is the process of creating a user-defined data type, and it is often referred to as "information binding". In object-oriented programming, abstraction is used to define object classes that closely resemble real objects, such as invoices and products.

Accelerator board—A printed circuit board added to a PC to increase its performance.

Access method—The technique or the program code in the operating system that provides input/output services. It defines where a group of data will be stored on a medium. By including the access method in the basic operating system, computer makers have made the programmer's job much simpler. In tape drives, the access method is straightforward—a block of data is placed sequentially after the last one. In disk drives, do not usually play data in sequential tracks. The access method software places a block of data in an available empty space and creates an index called a File Allocation Table (FAT), that rotates where the block of data can be found for retrieval later.

Adaptive compression—Data compression software that continuously analyzes and compensates its algorithm (technique) depending on the type and content of the data and the storage medium.

Address—Disks and other storage devices have numbers that identify locations by sector and by byte, like addresses on a city street. Retrieval software searches for the address assigned to the desired information in order to locate it.

Addressability—The ability to place information at a certain chosen area in an image.

Addressable capacity—The number of locations on an image that are addressable. To calculate, multiply the addressable vertical positions (row) by the addressable horizontal positions (column). Think of a matrix of dots, eight across by 16 down. The addressable capacity of the matrix is 128.

Agent—A piece of software that performs a role in a work task. Work tasks that routine and do not require the intervention of a person can be automated as part of the re-engineering effort.

Aggregation—Aggregation is an object-oriented technique that allows individual objects to be grouped together to form a metaobject that provides all of the interfaces or methods of its constituent objects.

Algorithm—Prescribed set of mathematical steps which is used to solve a problem or conduct an operation.

Aliasing—Condition when graphics, either constructed with lines (vectored) or dots (bitmapped), show jagged edges under magnification.

All Points Addressable (APA)—Refers to an array (bit-mapped screen, matrix, etc.) in which all bits or cells can be individually manipulated.

Allocate—To reserve the required amounts of a resource, such as disk space.

Alphanumeric—Set of characters composed of letters and numbers; may or may not include punctuation marks and other symbols; excludes printer control characters such as Carriage Return and flow control characters such as XON and XOFF.

American National Standards Institute (ANSI)—A standard-setting, non-governmental organization, which develops and publishes standards for "voluntary" use in the United States. Standards set by national organizations are accepted by vendors in that country.

American Standard Code for Information Interchange (ASCII)—The most popular coding method used by small computers for converting letters, numbers, punctuation and control codes into digital form. Once defined, ASCII characters can be recognized and understood by other computers and by communications devices. ASCII represents characters, numbers, punctuation marks or signals in seven on-off bits. A capital "C", for example, is 1000011 while a "3" is 0110011.

Analog—Comes from the word analogous, meaning "similar to." Analog devices record or monitor real world happenings, motion and sound, for instance, and convert them into "analogous" electronic representations, i.e. film or audio tape. Analog means recreating the continuous nature of the original thing. It is the opposite of digital, which translates the original happening into ones and zeros—an "unanalogous" representation.

Analog monitor—Video monitor that accepts an analog signal from the computer (digital to analog conversion is performed in the video controller). Analog monitors can be designed to accept a narrow range of display resolutions (for example, only VGA or VGA and Super VGA), or multisysnc analog monitors can accept a wide range of resolutions including TV (NTSC). Color monitors accept separate red, green and blue (RGB) signals for sharper contrast.

Annotation—The ability to attach notes to graphics or images. Useful for clarifying documents or editing images.

Anti-aliasing—Blending techniques that smooth the jagged edges of computer generated-graphics and type. A common anti-aliasing technique is to fill the pixels between the jagged ends with levels of gray (or color) to soften the edge and blend it smoothly into the background.

Anti-glare—An adjective used to describe the monitor screen that has been treated, coated or covered with a transparent substance that reduces the glare or reflection on the screen from office lights or sunlight. See OCLI.

Application—A broad and generic term for any software program that carries out a useful task. Word processors and graphics programs are applications.

Application Program Interface (API)—Generic term for any language and format used by one program to help it communicate with another program. Specifically, an imaging vendor can provide an API that enables programmers to repackage or recombine parts of the vendor's imaging system, or integrate the imaging

systems with other applications, or to customize the user interface to the imaging system.

Architecture—Refers to the way a system is designed and how the components are connected with each other. There are computer architectures, network architectures and software architectures.

Archival quality—The extent to which a reproduced image will (or won't) last "forever."

Archive—A copy of data on disks, CD-ROM, mag tape, etc., for long-term storage and later possible access. Archived files are often compressed to save storage space.

ASCII sort—A means of alphabetizing that accounts for capital letters and numbers. To arrange something in an ASCII sort, numbers (digits) come first in numerical order, followed by capital letters in alphabetical order, followed by lower case lower characters in alphabetical order.

Aspect ratio—The relationship of width to height. When an image is displayed on different screens or on paper or microform, the aspect ratio must be kept the same. Otherwise the image will be "stretched" either vertically or horizontally.

Association for Information and Image Management (AIIM)—Trade association and professional society for the micrographics, optical disk and electronic image management markets.

Asynchronous—Mode of data transmission in which each character is transmitted as a separate message often identified by start and stop bits.

Asynchronous join—Asynchronous join, or OR-Join, workflow routing provides asynchronous receiving of route information. In this type of join, route information from each of the nodes is processed individually and then continued on the specified route patch.

Attribute—In graphics, the condition a font is in— i.e. boldface, italic, underlined, reverse video, etc. is its attribute. In MS-DOS, files can be assigned attributes

that define how accessible it is, i.e., "read-only" is a file's attribute. In a document retrieval system, an attribute of a file is one of the keys by which the document has been stored and indexed.

Audit trail—Record of activity that has occurred in a certain file, or on a certain computer.

Authorization code—Identifying code, often a pass-word, that allows a user access to a system. Used mainly for privacy and security. Also used to divide up a comput-er's capacity among departments and/or hierarchies; dif-ferent "grades of service" are given different authoriza-tion codes.

Autochanger—A device that holds multiple optical discs and one or more disc drives, and can swap discs in and out of the drive as needed. Same as a jukebox.

Automated Tape Library (ATL)—Large-scale tape storage system, which uses multiple tape drives and mechanism to address 50 or more cassettes.

Automatic Document Feeder (ADF)—Scanner accessory that automatically feeds a stack of paper into the scanner.

Backbone—The part of the communications network which carries the heaviest traffic. The backbone is also that part of a network which joins LANs together—either inside of a building or across a city or the country. LANs are connected to the backbone via bridges and/or routers and the backbone serves as a communications highway for LAN-to-LAN traffic.

Backfile conversion—The process of scanning in, indexing and storing a large backlog of documents on an imaging system.

Background—1) The simultaneous, non-interrupting, execution of an automatic program while the computer is being used for something else. 2) The portion of micro-film that doesn't have anything recorded on it. It may be opaque or clear, depending on whether the film has a negative (background is opaque) or a positive (background is clear) image.

Background ink—A reflective ink used to print the parts of a document that are not meant to be picked up by a scanner or optical character reader.

Backlit—Any screen that has a light source which shines from the back of the image toward the viewer, making images sharper and easier to see in low ambient lighting conditions.

Backup—A duplicate copy of data placed in a separate, safe "place"—electronic storage, on a tape, on a disk in a vault—to guard against total loss in the event the original data somehow becomes inaccessible. Generally for short-term safety.

Bar code—A system of portraying data in a series of machine-readable lines of varying widths. The "UPC" on consumer items is a bar code. In document manage-ment, a bar code is used to encode indexing information. In microfiche, bar codes allow the automatic control of the duplication process, and contain indexing informa-tion. These bar codes usually appear in the last two or three title frames in the first title row of a microfiche.

Batch processing—Conducting a group of computer tasks at one time, instead of steadily throughout the day.

Baud—Unit of transmission speed equal to the number of signal events per second. In asynchronous transmis-sion, the unit of signaling speed corresponding to 1 unit interval per second; that is, if the duration of the unit interval is 20 milliseconds, the signaling speed is 50 baud. Technically baud is the same as "bits per second" when, and only when, each signal event represents exactly 1 bit (which is rarely true), but in casual, non-technical usage, baud is often misused to mean bits per second.

Bilevel—A binary scan that assigns each pixel an attri-bute of either black or white—no gray tones, no colors.

Binary code—Code which represents information as a sequential series of 0s and 1s.

Binary digit (bit)—Represents the binary code (0 or 1) with which the computer works. NOTE: The bit can take the form of a magnetized spot, an electronic

impulse, a positively charged magnetic core, etc. A number of bits together are used to represent a character in the computer.

Binary Large OBjects (BLOB)—the ability to embed large binary objects (images) as part of a character database record.

Bit—Contraction for Binary digit. The smallest unit of data a computer can process. Represents one of two conditions: on or off; 1 or 0, mark or space; something or nothing. Bits are arranged into groups of eight called bytes. A byte is the equivalent of one character.

Bit map—Representation of characters or graphics by individual pixels, or points of light, dark, or color, arranged in row (horizontal) and column (vertical) order. Each pixel is represented by either one bit (simple black and white) or up to 32 bits (fancy high definition color).

Bit-mapped image—Representation of image data where each pixel has a corresponding memory element.

Bit mapping—Creating rectangles over documents, mostly white=zeros and black=ones, about 1 million spots per page.

Bit specifications—Number of colors or levels of gray that can be displayed at one time. Controlled by the amount of memory in the computer's graphics controller cards. An 8-bit controller can display 256 colors or levels of gray. A 16-bit can show 64,000 colors. A 24-bit controller can display 16.8 million colors or gray levels.

Bit-mapped font—A set of dot patterns that represent all the letters, characters and digits in a type font at a particular size.

Bit-mapped graphics—Images which are created with sets of pixels, or dots. Also called raster graphics. Contrast with vector graphics.

Black and white scanner—Scanner that interprets scanned data as black or white, but with additional software, can perform electronic screening, dotting or dithering to produce simulated gray scale pixel configurations.

Black light—Term applied to radiant energy lying outside the visible range in the ultraviolet region of the spectrum. NOTE: It can be converted to visible light by the action of suitable fluorescent material.

Black line—A positive image, black on a clear or white background. Opposite of "white line." Also known as a negative image.

Block—The amount of data recorded contiguously on magnetic tape or disk in a single operation. Blocks are separated by physical gaps, or identified by their track/sector addresses.

Board—A fully functioning set of circuits installed on a circuit board and inserted into a computer system in order to provide additional processing functionality, typically not available from software programming alone.

Buffer—Device or allocated memory space used for temporary storage. Printers commonly use buffers, for example, to hold incoming text because the text arrives at a much faster rate than the printer can output.

Bug—An error in a computer program causing it to fail unexpectedly.

Burn-in—(1) Running a device for an extended period of time in order to ensure its functionality. (2) Also refers to the tendency of older computer displays to permanently display the "ghost" of a previous set of characters or graphics on the display because the image was held on screen for an extended period of time without changing it or dimming the display.

Bus—Signal path or line shared by many circuits or devices. Information is often sent to all devices throughout the same bus; only the device to which it is addressed will accept it. This makes designing system architecture much easier; devices can be plugged in "anywhere on the bus."

Business Process Automation—The use of computer-based information technology (specifically workflow technology) to automate the steps in a business process, coordinate the assignment and distribution of work items

and information among individuals, and manage the completion of tasks, activities and ultimately business processes.

Business Process Redesign (BPR)—The review, evaluation and redefinition of the tasks and activities that comprise a business process. The objective of BPR is to develop more efficient business processes.

Business Process Reeingineering (BPR)—The radical restructuring of the business processes, organizational boundaries, and management systems of an organization. Business process redesign and business process automation are components of BPR.

Byte—Eight bits of data grouped together to represent a character or some other computing data.

Cache—Small portion of high-speed memory used for temporary storage of frequently used data. Reduces the time it would take to access the data, since it no longer has to be retrieved from the disk.

Calibration—Process that adjusts color or grayscale values in an image for consistency among software applications and peripherals in scanning, displaying and transferring for color and black and white images.

Case—An individual instance of work to be performed for a business process; it can consist of one or more folders, documents and forms.

Case sensitive—Knows the difference between capital letters and lower case letters. A case-sensitive search for "CASE" would not find "case".

Catalog—Another name for a listing of directories or files stored on a computer or disk.

Cathode Ray Tube (CRT)—The glass, vacuum display device found in television sets and computer terminals.

Centralized processing—All or substantially all of an enterprise's computing is done in one site, usually called the data center. This was the norm in the US until the penetration of desktop PCs, which led the way to distributed processing.

Channel—Path or circuit along which information flows.

Character—A single letter, digit or punctuation symbol. A character equals a byte.

Character pitch—The number of characters per inch. The "tightness" of letters a printer can accomplish.

Character recognition—The ability of a machine to read human-readable text.

Character-Based User Interface (CUI)—Computer control system that makes the user type in commands (characters) to operate the computer. Opposite of GUI which uses pictures, or "icons," to help the user operate the computer. PCs running MS-DOS use CUI; Macs use GUI.

Characters Per Inch (CPI)—The density of characters per inch on tape or paper.

Charge-Coupled Device (CCD)—A type of digital camera technology in which the image is focused on an array of sensing pixels. The small size of the array itself, approximately microchip size, and the high resolution, around 1,000 and 1,018 pixels, of these cameras have greatly enhanced "image acquisition" capabilities and opened up exciting new applications in manufacturing quality control and in medicine.

Class—A class is an abstract data type defining the data and methods for a specific type of object. Programmers use classes to define instances of objects within their programs.

Class hierarchy—A class hierarchy is a group of superclasses that are related through an inheritance tree. In object-oriented programming, class hierarchies are used to identify common data and procedures once as a superclass, and then to enable the superclass to act as a template for related object types.

Client/Server—The relationship between machines in a communications network. The client is the requesting machine, the server the supplying machine. Also used to

describe the information management relationship between software components in a processing system.

Cluster—Group of terminals or workstations on the same system.

Common Object Model—The Common Object Model is the functional equivalent of the Component Object Model for UNIX-based platforms that today include SunOS, OBM, AIX, HP-UX, ULTRIX, OSF/1 and OpenVMS. The Common Object Model defines a common DCE RPC-based protocol and a subset of core OLE functions that Digital and other interested companies plan to support within their products.

Communications protocol—A set of communications rules that allows two devices to communicate with each other and check for potential errors to make sure transmitted data are not lost.

Compact Disk (CD)—A standard medium for storage of digital data in a machine-readable form, accessible with a laser-based reader. CDs are 4-3/4" in diameter. CDs are faster and more accurate than magnetic tape for data storage. Faster, because even though data is generally written on a CD contiguously within each track, the tracks themselves are directly accessible. This means the tracks can be accessed and played back in any order. More accurate, because data is recorded directly into binary code; mag tape requires data to be translated into analog form. Also, extraneous noise (tape hiss) associated with mag tape is absent from CDs.

Compact Disk Read Only Memory (CD-ROM)—A data storage system using CDs as the medium. CD-ROMs hold more than 600 megabytes of data.

Component inheritance—Component inheritance allows OLE component objects to be easily reused in different applications, without creating implicit relationships between objects. Component inheritance is the use of an existing component object to supply functionality for a new object.

Component Object Model—The Component Object Model, or COM, is a standard mechanism for objects written by different companies in different programming languages to interact. COM is the basic "wiring and plumbing" for all OLE features. For instance, COM allows component software applications to be integrated into larger business systems through OLE Automation.

Component software—Component software is an application that contains one or more component objects that can freely interact with other component software through OLE capabilities.

Composer—The route composer is the person who originates the workflow route.

Compound documents—Compound documents are documents which contain multiple data types. Often, the different types of data have been created by different applications, and embedded into the document.

Compression—A software or hardware process that "shrinks" images so they occupy less storage space, and can be transmitted faster and easier. Generally accomplished by removing the bits that define blank spaces and other redundant data, and replacing them with a smaller algorithm that represents the removed bits.

Computer Output to Laser Disk (COLD)—Technique used to transfer computer-generated output to optical disk.

Computer Output Microform (COM)—The process of converting data (having been input by a number of means) to microfilm or microfiche.

Computer readable—Data which is in a format, such as ASCII, or on a medium, such as disks, tapes, optical discs or punched cards, that a computer can understand.

Conditional routing—Conditional routing enables the route composer to build conditional statements into a workflow route.

Constant Angular Velocity (CAV)—Technique enabling data recorded with a variable linear density to be read, whereby the speed of rotation of the disk remains constant.

Constant Linear Velocity (CLV)—The technique of adjusting the speed of a disc's spinning, so that the larger outer tracks (which normally would spin faster) can be slowed down and thus hold more data than the smaller inner tracks. Used in CD-ROM.

Consultative Committee for International Telegraph and Telephone (CCITT)—International organization that develops international communications standards.

Container—A container application is an OLE-enabled application that can store embedded or linked objects provided by OLE server applications. When a user drags a spreadsheet chart into a word processing document, for example, the spreadsheet application is the OLE server and the word processor is the OLE container. OLE-enabled applications can be both object containers and object servers.

Contention—When two or more users try to access the same device at the same time. There are "collision prevention" techniques to solve contention problems in LANs.

Contextual search—To locate documents stored in a system by searching for text that appears in them, rather than by searching for them by file name or other indexing technique.

Contiguous—Placed adjacently; one after another.

Controller—A hardware/software device that facilitates communications between a host and one or more devices.

Copy—A duplicate of the original. A digital copy (from CD to CD for instance) will be perfectly identical. The condition is binary, the signal is either on or off, no "noise" in between. An analog copy (from mag tape to mag tape, for instance) will likely degrade each time a copy is made (called generations) because of tape noise.

CORBA—Common Object Request Broker Architecture. A standard which defines the manner in which software "objects" created in one program can be used in another.

Cursor—The symbol on a screen that shows where the next activity will take place. Graphics programs often change the shape of the cursor, depending on what action the computer is programmed to take next.

DAT Auto Loader (DAL)—Device that accepts a magazine of five or so DAT tapes, which are each addressable.

Data compression—Reducing the amount of electronic "space" data takes up. Methods include replacing blank spaces with a character count, or replacing redundant data with shorter stand-in "codes." No matter how data is compressed, it must be decompressed before it can be used.

Data decompression—The regeneration of a bit-map from a compressed representation.

Data transfer—The movement of data inside a computer system.

Database—Data that has been organized and structured in a disciplined fashion, so that access to information of interest is as quick as possible. Database management programs form the foundation for most document storage indexing systems.

Database Management System (DBMS)—Set of programs designed to organize, store and retrieve machine-readable information from a computer-maintained database or data bank.

Decompress—To reverse the procedure conducted by compression software, and thereby return compressed data to its original size and condition.

Degausser—A device that removes unwanted magnetism from monitors or the heads in a tape or disk drive mechanism.

Delimiter—The "divider" character, often a comma, between separate fields in database records.

Demodulation—Extracts the information (digital or analog) from the carrier signal, so that the transmitted information may be used.

Descriptor—The key word, code or phrase that an automated document retrieval system uses to identify and locate the document. Descriptors sometimes "summarize" the most relevant data in the document, so that reading the descriptors—rather than retrieving the entire document—is sometimes sufficient for the purposes of the search.

Desktop—Slang for any computer function that can be done on a standalone PC, rather than a larger, more powerful, computer.

Device drivers—Programs that tell the computer how to communicate with particular peripheral devices.

Digital—The use of binary code to record information. "Information" can be text in a binary code like ASCII, or scanned images in a bit mapped form, or sound in a sampled digital form, or video. Recording information digitally has many advantages over its analog counterpart, mainly ease in manipulation and accuracy in transmission.

Digital Audio Tape (DAT)—A technology that records noise-free digital data on magnetic tape. Generally used for audio, a DAT cassette can hold 0 to two gigabytes when adapted for data storage.

Digital camera—The newest generation of video cameras transform visual information (lightness and darkness) into pixels, then translate the pixel's level of light into a number (or, in the case of color, into three numbers—one for the level of red, green and blue in the pixel). These digital images can then be manipulated pixel by pixel to create exciting new applications in video and film production. They can also be compressed, stored and transmitted in more or less the same manner as traditional digital data.

Digital Data Storage (DDS)—A DAT format for storing data. It is sequential; all data that is recorded to the tape falls after the previous block of data.

Digital image—Image composed of discrete pixels of digitally quantized brightness.

Digital scanner—Optical reader that scans and converts images into digital form.

Digitization—Use of a scanner to convert documents (on paper or microforms) to digitally coded electronic images suitable for magnetic or optical storage.

Digitize—To convert an image or signal into binary code. Visual images are digitized by scanning them and assigning a binary code to the resulting vector or raster graphics data. Sounds are digitized by recording frequent "samples" of the analog wave, and translating that data into binary code.

Digitizer—Device for the digitization of a document. This term is often used, by extension to refer to a device that allows both the scanning and the actual digitization of the document.

Direct Access Storage Device (DASD)—Any on-line data storage device. A disc, drive or CD-ROM player that can be addressed is a DASD.

Disc—A digital storage medium. Optical discs are made of a metal alloy recording surface sandwiched between a rigid substrate and a plastic protective coating. Lasers record data in the metal alloy by either creating tiny pits (ablation technique) or by causing small bubbles to form in the "negative" area, thereby reflecting the laser away. Generally, disk with a "c" means optical disc. Disk with a "k" means magnetic hard or floppy disk.

Disk—A round, flat recording medium which consists of a substrate(s) with one or more layers deposited on the surface(s) onto which information can be recorded and played back when the disk is loaded in a disk drive.

Disk array/Disc array—Combining redundant disk or disc drives for more capacity, or for disaster recovery.

Disk array controller—Acts as a manager between the host and the drives. Comprised of a main computer module, channels for each drive, and a host channel for each host input. Adding memory modules increases performance.

Disk drive—A device containing motors, electronics and other gadgetry for storing (writing) and retrieving (reading) data on a disk. A hard disk drive is generally not removable from the machine. A floppy disk drive accepts the removable disk cartridges.

Disk duplexing—A method of fail-safe protection, occasionally used on file servers on local area networks. Disk duplexing involves copying data onto two hard disks simultaneously, each through a separate disk channel. The idea is, if one disk or channel is faulty, the other will most likely continue to operate normally.

Disk management—Refers to the control of information stored on a disk. The logical relationship of subdirectories to root directories, for instance.

Disk mirroring—A fault-tolerant technique that writes data simultaneously to two hard disks using the same hard disk controller. The disks operate in tandem, constantly storing and updating the same files. Mirroring alone does not ensure data protection. If both hard disks fail at the same time, you will lose data.

Disk pack—A cartridge of hard disk platters arranged as a single unit. A disk pack contains more space for storing and retrieving information than one single disk.

Disk (file) server—A mass storage device that can be accessed by several computers, usually through a local area network (LAN).

Disk sector—Magnetic disks are typically divided into tracks, each of which contains a number of sectors. A sector typically contains a predetermined amount of data, such as 256 bytes.

Disk striping—Spreading data over multiple disk drives. Data is interleaved by bytes or by sectors across the drives.

Display—Commonly used to refer to the device utilized for viewing images and data in a computing environment. Also used to refer to the visual presentation of data.

Dithering—Simulating gray tones by altering the size, arrangement or shape of background dots.

Document—A collection of data, organized into some logical order and presented on paper.

Document preparation—Steps to ready documents for filming or scanning, e.g., removing paper clips, staples, bindings and sorting by categories.

Document retrieval—The ability to search for, select and display a document or its facsimile from storage.

Dots per inch (dpi)—A measurement of resolution and quality. Measures the number of dots a printer can print per inch both horizontally and vertically. A 600 dpi printer can print 360,000 (600 by 600) dots on one square inch of paper. More dpi means higher resolution and greater detail.

Dot pitch—The distance of one phosphor dot in a CRT to the nearest phosphor dot of the same color on the adjacent line.

Drum—Cylinder in a scanner that "holds" the original document during the scanning process.

Duplex—(1) In micrographics, a method of recording on roll microfilm in one exposure the images of the front and back of a document. The microimages appear side by side across the width of the microfilm (ISO). (2) Term applied to any scanner capable of performing duplex work as described in (1). (3) In communications, the ability to send and receive information simultaneously.

Dynamic Data Exchange (DDE)—Allows one free standing program to give commands, take requests, and give and receive data from another free standing program in a Microsoft Windows environment.

Dynamic Link Library (DLL)—The Microsoft Windows specification for linking program subroutines from one application with the subroutines of another.

Easy scale—A Windows DLL that provides high quality scale to gray and N:M scaling for any bitonal image on all monitors when implemented in the viewing software.

Electron gun—The device in the CRT that produces the electron bream that activates the phosphors, causing them to emit red, green and blue light.

Electronic Data Interchange (EDI)—An electronic communications standard which connects business trading partners for conducting contract negotiations, sales, invoicing and collections.

Electronic forms—Graphics that are merged electronically with data. Can be as simple as a borderbox, or a logo or running header.

Electronic image—Digital representation of a document.

Electronic image gray scaling—Activity outside or in scanning that accurately senses, differentiates and encodes intermediate shades between black and white in photographs and half tones.

Electronic imaging—Electronic techniques for capturing, recording, processing, storing, transferring and using images.

Electronic mail—A means of connecting computers in order to send messages to one or more individuals or groups.

Emulation—Imitation of a function by a system not originally designed to perform that function.

Encapsulation—A technique used to isolate some of the decisions made in writing a program. To encapsulate decisions a program is organized into an interface, such as a set of procedures and an internal part. All access to the programs services are available only through the interface. The internal part of the program can be used, but not be accessed.

End of File (EOF)—Special character that marks the end of a file or other document. Used in both stored and transmitted data.

Enhanced Graphics Adapter (EGA)—A display technology for the IBM PC. It has been replaced by VGA.

Enhancement—Technique for processing an image so that the result is visually clearer than the original image.

Error Detection and Correction (EDAC)—An error detection scheme. It works this way: An pre-arranged extra block of data is added to each block written. After writing, the extra data is read back. If the extra data is correct, EDAC assumes the entire writing procedure went smoothly, and goes on to repeat the procedure with the next block. If an error is detected, the write process is repeated.

Error rate—Ratio of the amount of erroneously recorded, read or transmitted information to the total amount sent.

Ethernet—Particular implementation of a bus-type local area network that communicates at 10 megabits per second.

Facsimile—A collection of technologies. Facsimile first scans, then digitizes a paper document. It then converts that digital image to analog form. The fax machine then dials and arranges a data communications session, agreeing on speed of transmission and protocol, with a remote machine. The analog version of the document is then transmitted. Meanwhile, the receiving machine captures the analog data, reconverts it to digital form and finally prints a copy, or facsimile, of the original document. Facsimile technology is in its third generation, called Group 3. Each step up has been an improvement in speed of transmission and resolution. Group 3 transmits a page at 9600 baud in less than a minute. Group 3 resolution is 203 x 98 dpi in standard mode and 203 x 96 dpi in fine mode. The standards for Group 4 facsimile exist, but its use is restricted to private corporate application at present, since it requires an entirely digital transmission network. When the public telephone network is all-digital, general-use Group 4 fax machines will become commonplace.

Facsimile transmission—Process by which a document is scanned, converted into electrical signals, transmitted, and recorded or displayed as a copy of the original.

Fault tolerant components—Fault tolerance implies that if any component of the subsystem fails, the unit will remain operational.

Fax board—An add-on circuit board, that fits in a PC, that sends computer files in fax format to either a fax machine or another fax-board equipped PC. Quality of the image is better, since it isn't scanned. And when the output is directed to a laser printer, the image is even sharper and comes out on regular paper (as opposed to the dubious quality of fax machine's thermal paper).

Feature extraction—A sophisticated optical character recognition technique. The software keeps data regarding all characters' features, i.e., the letter "A" has two diagonal lines that intersect at the top; it has a horizontal line that crosses from one of the lines to the other, etc. As the OCR scans, it compares features of the character to its feature library. Feature extraction is used to recognize handwriting, in certain constrained cases.

Field—The smallest logically distinguished unit of data in a record. In a database, the individual items of related information, for example, policyholder's name, address, social security number, etc. "Logically distinguished" means that there are similar units of data in other records that have something in common. For example, "last name" is a field, an entire mailing address is a record. All of the address records is a database.

Field separator—The prearranged code, typically a comma, that separates fields in a record. Also called a delimiter: "The records in that database are comma-delimited."

File—All the data that describes one document or image, maintained under a single naming code and stored in a computer or in a storage medium.

File Allocation Table (FAT)—Data written to a magnetic disk is not necessarily placed in contiguous tracks. It is usually divided into many clusters of data in many locations on the disk surface. The FAT is the special area on a disk which keeps track of where clusters of data have been written for retrieval later.

File server—Local Area Network (LANs) were invented to allow users on the LAN to share and thereby conserve the cost of peripherals (printers, modems, scanners) and to likewise share software. The file server is the machine on the LAN where the shared software is stored.

File-oriented backup—Any backup software which instructs the computer to store information in files just as they appear on the originating computer, making restoration easier and more logical.

First In, First Out (FIFO)—Queue handling method that operates on a first-come, first-served basis.

Fixed disk—Another name for hard disk. So-called because it is installed in a computer and not meant to be removed.

Flat-bed scanner—Device for scanning that has a flat surface for input material. Generally used for scanning bound material.

Flat profile screen—A screen that appears almost entirely flat or with little to no convexity when viewed from the side. These screens reduce reflection, glare, and distortion that that may occur as information is displayed closer to the corners of the screen.

Flip—The technological equivalent of the turn of a page.

Flowchart—A diagram that uses symbols and interconnecting lines to show the logic and sequence of specific program operations. Also used to show the sequence and logic of processing to achieve objectives

Focus servo—The device in an optical drive that keeps the read/write beam aligned despite imperfections in the medium or bumps and shakes

Folder—(1) A term for the basic element in its file management scheme. A folder holds sets of directories. A folder can hold other folders. It is basically a hierarchical tree-directory scheme, just like DOS's directories and sub-directories. (2) A logical collection of electronic documents stored on the document management system.

Font—All the characters and digits in the same style and size of type.

Footprint—The physical area a machine occupies on your desk; the amount of square feet, or "real estate," devoted to a machine.

Form—The added-on border of a microform, which may consist of a simple border or may contain identifying matter such as logos or titles.

Form factor—The size of a mechanism, usually refers to units meant to be installed in a PC or workstation. If a new hard drive is six inches deep, and your PC can only accept five inches, then you say its "form factor" is too large.

Formatted data—Data which has been processed with software to attach the necessary titling, indexing, and job separation instructions.

Formatting—Preparation of a storage medium, defining tracks, check for bad sectors, etc.

Full text search—The ability to search text files for occurrences of certain words digits, sentences or patterns of characters. Generally, a scanned document cannot be full text searched. To do that, the document would have to be retyped or scanned with an OCR to create a text file.

Gesture recognition—The ability to electronically recognize handwritten characters, check marks and certain other symbols.

Giga—Meaning billion or thousand million. In computers, it is actually 1,024 times mega and is actually 1,073,741,824. One thousand gigas is a tera.

Gigabyte (GB)—A million (actually more) bytes of data, or a thousand megabytes. Imaging applications commonly take up huge amounts of data. For example, it only takes ten 8 1/2 x 11-inch color pictures, scanned at 600 dpi, to fill a gigabyte.

Global hot spare—A spare drive that is continuously powered up and spinning. In the event that any drive fails, this spare drive replaces the failed drive straight away, providing the array with immediate access to a functioning drive. Allows the system to recover from a drive failure anywhere in the array before the failed drive is replaced.

Graphical User Interface (GUI)—Computer control system that allows the user to command the computer by "pointing-and-clicking," usually with a mouse, to pictures or "icons," rather than typing in commands.

Gray scale—The spectrum, or range of shades of black an image has. Scanners' and terminals' gray scales are determined by the number of gray shades, or steps, they can recognize and reproduce. A scanner that can only see a gray sale of 16 will not produce as accurate an image as one that distinguishes a gray scale of 256.

Halftone—A graphic, usually created from a photograph, in which dots are used to represent continuous tones. Larger, densely placed dots which sometimes touch represent darker tones; smaller, widely spaced dots with white areas between them represent light tones. Color halftones use varying hues and combinations of the subtractive, or "process," colors to represent full continuous tone images. Halftones allow continuous tone photographs to be printed by conventional ink-on-paper processes. There is at present no way to print actual continuous tones.

Handprint Character Recognition (HCR)—The ability of a computer to read handprinted characters, not generated by machine.

Handshaking—Exchange of signals at the beginning of a data communications session. During this exchange, the two systems confirm each other's specs like parity, baud rate and speed, to ensure a proper link is set for the data transmission. As with humans once the handshaking is through, the business of communications begin.

Hard disk—A storage device that uses a magnetic recording material. Generally, hard disks are fixed inside a PC, but there are removable cartridges versions. Hard disks store anywhere from five to hundreds of megabytes.

Header sheet—An instruction sheet for an optical character reader that defines the format of the pages to be scanned.

Hierarchical File System (HFS)—In DOS, the file management system that allows directories to have subdirectories, and sub-subdirectories. In Macintoshes files may be placed into folders, and folders to be placed within other folders.

High Resolution (Hi-res)—Basically, any image that is displayed in better quality by increasing the number of dots, or pixels per inch than normal. Usually refers to better quality computer displays, but can describe printer quality as well.

Horizontal scan frequency—The number of video lines written on the screen every second (left to right). The higher the horizontal scan frequency, the higher the resolution and/or the refresh rate.

Horizontal table—In indexing, a table with entries that follow one another sequentially, i.e. entry number one is byte number one; entry two is byte two.

Hot redundancy—A component or system runs parallel with an identical "twin." Should one twin fail, the other is already running and provides full service without interruption.

Hot swap—A method used to remove or replace hardware system components while the system remains powered on and operating. Subsystem activity is not significantly affected while the host swap is in progress.

Icon—The basis of a graphical user interface. An icon is a picture or drawing of a device or program which is activated, usually with a mouse, to access the device or run the program.

Image—Digital representation of a document, picture or graphic.

Image processing—Refers to the manipulation of raw data to solve some problems or enlighten the user in some way not possible without the manipulation. Digitized images which have been "acquired" (scanned, captured by digital cameras) can be manipulated. The purpose may be simply to improve the image, to change its size, color, or simply to touch-up parts of it. A more important application of image processing is to compare and analyze images for characteristics that a human eye alone couldn't perceive. This ability to perceive minute variations in color, shape and relationship has opened up applications for image processing in high-speed manufacturing, quality control, criminal forensics, medicine, defense, entertainment and the graphic arts.

Image processor—Device that takes input data and changes it into the proper format for an imaging device—printer, display, microform, or computer.

Image resolution—The fineness or coarseness of an image as it was digitized, measured as dots-per-inch (dpi), typically from 200 to 400 dpi.

Imaging—Recording "human-readable" image—pictures, images, motion, text, etc., into "machine-readable" formats, i.e. microfilm, computer data, videotape, OCR output, ASCII code, etc.

Imaging system—Collection of units that work together to capture and recreate images. At its simplest, it has an acquisition device (scanner, camera), an image processor and an imaging device (printer, microfilm, computer).

Incremental backup—Backing up only files that have been changed since the last backup, rather than backing up everything.

Index—A descriptive set of data associated with a document for locating the document's storage location. In a more complex and demanding role, indexing can be used to consolidate documents that may not be, at first glance, related, or that may be stored in different locations, or on different media. Indexing stored documents is the great intellectual challenge in document retrieval. Anyone can scan a piece of paper to microfilm. The hard part is devising an indexing scheme that describes every possible parameter of each document for later searches, comparisons and processing.

Indexing—A method by which a series of attributes are used to uniquely define an imaged document so that it may later be identified and retrieved.

Inheritance—Is a mechanism for sharing code and behavior. It allows for the reuse of the behavior of a class in the definition of new classes.

Initialize—Startup process in which a device or system is prepared, or automatically prepares itself, for normal operation. Usually returns all parameters to their default values.

Input—(1) Process of entering information into a system, e.g., a computer. (2) Data entered into a computer.

Input/Output (I/O)—Refers to the process, techniques and media used for human/machine communications. Also refers to data submitted between computing components.

Intelligent Character Recognition (ICR)—Advanced form of OCR technology that may include capabilities such as learning fonts during processing or using context to strengthen probabilities of correct recognition.

Interface—An interface is simply a mechanism for different pieces of software to interact. For instance, application programming interfaces (APIs) are provided with operating systems to access system-level services from programming languages; database management systems to access SQL database services; and any number of other types of applications and system software.

Interlaced—Only every other line of pixels on a TV or computer terminal screen is refreshed on each "pass" (in American television, which is interlaced, every second line is refreshed 60 times a second). Interlacing thereby saves half the signal information that non-interlaced screens use.

Interleaved—The system of writing to a hard disk that places data in non-contiguous tracks because of the rapidly spinning nature of a disk drive. The operating system keeps a "log" of where each sector of data is stored for retrieval later.

Jaggies—Slang for aliasing. The ragged or stair-stepped appearance of diagonal lines and curves.

Jam—Paper misfeed in a scanning device

Jitter—The flickering of a displayed image. Sometimes the result of interlacing.

Joint Bi-level Image Group (JBIG)—Algorithm standard for bi-tonal compression which has applications in database management systems that are composed of black and white half-toned photos and text.

Joint Photographic Experts Group (JPEG)—Proposed standard for still image compression. Devised by the Joint Photographic Experts Group, sanctioned by the International Standards Organization (ISO) and the CCITT. A color image is digitized into pixels, each with a numerical value that represents brightness and color. The picture is then broken down into blocks, each 16 pixels by 16 pixels and then reduced to 8 pixels by 8 pixels by subtracting every other pixel. The software uses a formula that computes an average value for each block, permitting it to be represented with less data. Further steps subtract even more information from the image. To retrieve the data, the process is simply reversed to decompress the image. A specialized chip decompresses the images hundreds of times faster than is possible on a standard desktop computer.

Jukebox—A device that holds multiple optical discs and one or more disc drive, and can swap discs in and out of the drive as needed. Same as an autochanger.

Key—A word, number or phrase associated with a document to aid in its retrieval from storage. Sometimes called descriptors. There are often many keys used together to fully locate a document; together they are called an index.

Keyword—A word associated with a document or document image to aid in its retrieval from storage.

Kilobyte (Kbyte)—One thousand bytes. To a computer, it is actually 1,024. So 16 Kbytes, or 16K is actually 16,384 bytes, 64K is 65,536, etc.

Land and groove—A physical feature of optical discs, applied during manufacture, which defines track locations. The groove is recordable, the land separates the grooves and is not recordable.

Landscape—Page or monitor orientation in which the page width exceeds the page length. Also called "comic" after the shape of frames in comic strips.

Laser—Source that produces light that is nearly monochromatic (of only one wavelength) and highly coherent (with waves in phase both temporally and spatially). Acronym for Light Amplification by Stimulated Emission of Radiation.

Laser disc—An optical disc with the same technology as a Compact Disc, except laser discs are 1" in diameter.

Laser fax—A conventional laser printer that also can be used as a fax machine when combined with an optional plug-in cartridge and used with a personal computer.

Laser optical—System of recording on grooveless discs using a laser-optical-tracking pickup.

Laser printer—Printer that uses a beam of light to charge a drum so that it attracts toner, which is transferred to heated paper.

Last In, First Out (**LIFO**)—A queuing or inventory scheme whereby the most recent "thing" to come in is acted on first.

Leading—The space between lines of printed text. OCRs have to be adjusted for the leading of a document to read it properly.

Lens—Converging optical system consisting of refracting components designed to form real optical images which may be recorded on a sensitive surface or viewed on a screen (ISO).

Line Screen—The resolution of a halftone, expressed in lines per inch. Usually between 53 lpi and 150 lpi.

Line Segment—In vector graphics, same as vector.

Lines per inch (**lpi**)— Number of scanning or recording lines per unit length measured perpendicular to the direction of scanning.

Lines per minute (**lpm**)—One of the parameters by which electronic printers and scanners are judged.

Local Area Network (**LAN**)—Data communication network of connected devices within a small area, such as a building or group of buildings. High-speed transmissions over twisted pair, coax, or fiber optic cables that connect terminals, personal computers, mainframe computers, and peripherals together at distances of about 1 mile or less.

Logical—A feature that is not physically present, but applied by software. Sectors on a hard disk are physically arranged contiguously; logically, sectors may be placed anywhere on a hard disk.

Lossless—Image and data compression, applications and algorithms, such as Huffman Encoding, that reduce the number of bits a picture would normally take up without losing any data.

Lossy—Methods of image compression, such as JPEG, that reduce the size of an image by disregarding some pictorial information.

Low resolution (**Lo-res**)—Low quality reproduction because of a small number of dots or lines per inch.

Machine readable—Data which is in a format, such as ASCII, or on a medium, such as disks, tapes, optical discs or punched cards, that a computer can understand.

Magnetic ink—Ink that can be read by a magnetic scanner used on bank checks.

Magnetic Ink Character Recognition (**MICR**)— The ability, by a scanning machine, to recognize characters printed with magnetic ink. Used on checks to help banks sort them.

Magnetic media—Refers to the type of media used for storing data. Magnetic media utilizes metallic material capable of holding a magnetic charge to record information.

Magnetic recording—A technique of recording analog or digital signals or data on a medium of specially prepared grains of iron oxide; old-fashioned tape recording (although floppy and hard disks use basically the same technology).

Magnetic tape—Storage medium that uses a thin plastic ribbon coated with iron oxide compound to record data with electrical pulses. Mag tape is a sequential storage medium, the next bit of data is recorded after the last bit. In order to locate a specific bit of data, you have to look through the whole tape until you find it. The standard for data recording is nine-track mag tap; one byte (eight bits plus a parity bit) fits across the tape widthwise.

Magneto Optic Recording—Recording data using optical means to change the polarity of a magnetic field in the recording medium. Data is erasable and/or rewritable.

Management Information System (MIS)—Management information system that is provided by computer processing.

MAPI (Messaging API)—A Microsoft published API that separates the client from the server functionality, allowing various clients, like mail front ends, word processors, spreadsheets, etc, to access the messaging capabilities of back-end mail servers, such as Microsoft Exchange Server.

MAPI Workflow Framework—The MAPI Workflow Framework defines custom message classes and associated collections of properties to enable the interoperability of different workflow systems. By sending and processing MAPI Workflow command messages, workflow systems can interact and share information.

Mark—Same as blip. Small character printed or notched on microfilm for timing or counting purposes. On an optical disc, it may take the form of a pit, hold, bubble or light-reflective area.

Mark geometry—The size and shape of the mark made by a laser on an optical medium.

Mean Time Between Failures (MTBF)—A measure of equipment reliability (the higher the MTBF, the more reliable the equipment).

Mean Time To Repair (MTTR)—A measure of the complexity an modularity of equipment (the higher the MTTR, the more complex—or less modular—the equipment).

Media—Materials used to store information and data.

Megabyte (MB)—Approximately one million bytes. Precisely 1,024 kilobytes, or 1,048,576 bytes.

Megahertz (Mhz)—One million cycles per second.

Memory—Area of a computer system that accepts, holds, and provides access to information and data

Menu—A displayed set of options for the user in an interactive system. For instance, the list of relevant documents available after a search command has been completed.

Microfiche—Microform in the shape of a rectangular sheet having one or more microimages usually arranged in a grid pattern, with a heading area across the top.

Microfiche scanner—Device for scanning microfiche.

Microfilm—A film medium, in tape-like roll, for recording reduced pages of documents sequentially.

Micrographics—The branch of science and technology concerned with the methods and technique for recording information on, and retrieving it from, microform. Those methods include reducing and recording images by photographic means, or directly onto film by computer (computer output microform, or COM); the location and retrieval of documents through indexing and mechanical means; and the display and magnification on display screens or paper output.

Migrate—To move files from one storage medium to another, from on-line to near-line or near-line to off-line. Usually files are migrated when they match parameters set by network managers. These parameters include age, time since last access and size.

Mil—One one-thousandth (1/1,000) of an inch; used to describe paper and tape thickness.

Millions of Instructions Per Second (MIPS)—A measure of computer speed.

Millisecond—One thousandth of a second. Expressed numerically as 0.001 and abbreviated as ms. Or msec.

Mirror image—(1) Characterizing a reversal of orientation, as the image of an object formed by a plane reflecting surface. Right-to-left change, as seen in a flat mirror. (2) The reversed image of an object as seen in a mirror. See also reverse reading and right reading.

Mirroring—An exact copy of the contents of one disk drive on another. In the event of a disk drive failure, the second drive is accessed for the desired data. Also know as shadowing.

Modem—Device that allows digital signals to be transmitted and received over analog telephone lines. Short for modulator-demodulator.

Modified Constant Angular Velocity (MCAV)—Qualifies an optical disk which includes several centric zones. Each zone is read at constant angular velocity, crossing from one zone to another involves modifying the velocity.

Modified Constant Linear Velocity (MCLV)—Describes a disk on which tracks are divided into bands. Within a band, the disk spins at constant angular velocity, but that velocity is different for each band. The relation between velocity and band location is similar to velocity radius curve for CLV operation.

Moire—The undesired effect caused by overlaying dot patterns (usually halftone representation of photographs) which are incompatible.

Motion Picture Experts Group (MPEG)—An image-compression scheme for full motion video proposed by the Motion Picture Experts Group, an ISO-sanctioned group. MPEG takes advantage of the fact that full motion video is made up of many successive frames, often consisting of large areas that don't change, like blue sky background. MPEG performs "differencing," noting differences between consecutive frames. If two consecutive frames are identical, the second doesn't need to be stored.

Mount—Extend the hierarchy of directories. Accomplished over a LAN by associating the root of one computer with a directory of a mounted computer.

Mouse—Hand-driven input and pointing device for personal computers.

MS-DOS—The basic command system, call disk operating system (DOS), for IBM and IBM clone personal computers.

MTBF—Mean time between failure. Usually measured in hours.

Multimedia—Combining more than one media for the dissemination of information, i.e., using text, audio, graphics, animation and full-motion video all together. Requires enormous amounts of bandwidth and processing power.

Multisync monitor—A monitor that adjusts to the type of video signal it receives. MultiSync is trademarked to NEC.

Near-line—Data that is available on a secondary storage device that the user can access, but at a slower rate the on-line data is accessed.

Network topology—Physical arrangement of nodes and interconnecting communications links in networks based on application requirements and geographical distribution of users.

Node—A point of connection into a network. In multi-point networks, it means it is a unit that is polled. In LANs, it is a device on the ring. In packet switched networks, it is one of the many packet switches which form the network's backbone.

Object-Oriented Program (OOP)—Programming that views programs as a collection of autonomous

agents called objects. Each object is responsible for specific tasks.

Object-Oriented Programming System (OOPS)—Object-oriented programming, in contrast with procedural programming, involves the use of both object-oriented design and an object-oriented programming language such as C++ or Smalltalk.

Object technology—Object technology is a broad term that refers to the use of "objects" to (1) analyze; (2) model or design; and/or (3) implement some aspect of a computer system.

OCLI (Optical Coating Laboratories, Inc.)—The primary manufacturer of anti-glare treatments for monitor screens. The company name has become synonymous with its product.

Off-line—Data that is not physically stored on an accessible drive, such as removable tapes or disks.

OLE—OLE is a set of system services that provides a means for applications to interact and interoperate. Through OLE Automation, an application can dynamically identify and use the services of other applications. Applications that accept objects from other applications are called containers, while the application providing the object is called a server. Through OLE object linking, objects created in one application can be linked into container applications. As the linked object is changed or revised by the server application, it is automatically updated in any container applications. Through OLE object embedding the container application does not maintain a link to the object's data source, so updates to an embedded object must be made from within the document itself. Through OLE Visual Editing, embedded and linked objects can be directly edited within the container application without switching to the server applications.

OLE controls—OLE Controls are a special form of component Automation Object. OLE Controls are similar to Visual Basic custom controls (VBXes), but their architecture is based on OLE. This means that OLE

Controls can be freely plugged into any OLE-enabled development tool or application.

Omnifont recognition—The ability of an optical character reader to recognize a typeface font without having to "learn" (make a template in advance) that typeface. Omnifont character recognition uses feature extraction techniques.

On-line—Data that is available on a primary storage device so that it is readily accessible to the user.

On-line spare drive—Remains idle until one of the drives in the array fails, at which time it replaces the failed drive. The data that was on the failed drive is then recreated on the on-line spare utilizing the data and or parity values on the remaining two drives.

OpenDoc—OpenDoc is a specification for a compound document architecture that is being formed by the joining of several different technologies supplied by Apple (the base OpenDoc architecture, the Bento file system and the Open Scripting Architecture) and IBM (the System Object Model). The development effort for combining these technologies has been divided among key consortia members, including Apple, IBM, WordPerfect and Novell. To date, the OpenDoc software is not available on any platform. Unlike OLE, OpenDoc does not allow the integration of shrink-wrapped applications and components. See also the System Object Model/Distributed System Object Model.

Operating system—Collection of programs that, taken together, manage the hardware and software; it is the operating system that makes the hardware usable, providing the mechanisms that application programs use to interact with the computer.

Optical—(1) Containing lenses mirrors, etc., as in optical view-finder and optical printer. (2) In general, having to do with light and its behavior and control, as in optical properties, optical rotation. (3) Pertaining to the science of light and vision.

Optical Character Recognition Reader (OCR)—
The ability of a scanner with the proper software to capture, recognize and translate printed alphanumeric characters into machine readable text. Most OCRs work by using either Pattern Matching or Feature Extraction. With pattern matching, the software is given a "template" of possible characters. When the scanner sees a letter, it compares it to its library of pattern templates. If there is enough of a match, it safely assumes it has "recognized" the letter and sends the ASCII equivalent of the letter to the output file. Feature extraction is more sophisticated. Its "library" consists of groups of information regarding a character's features; i.e. the letter "A" has two diagonal lines; the lines intersect at the top; it has a horizontal line that crosses from one of the lines to the other, etc. As the OCR scans, it compares features of the character to its feature library. Feature extraction is used to recognize handwriting in certain constrained cases. All OCR software further supports its "guesses" by knowing a little something about the language. A digit "1" is not likely to fall in between a group of letters; the letter "h" frequently follows the letter "t," etc.

Optical disc—A direct access storage device that is written and read by laser light. Certain optical discs are considered Write Once Read Many (WORM), because data is permanently engraved in the disc's surface either by gouging pits (ablation); or by causing the nonimage area to bubble, reflecting light away from the reading head. Erasable optical drives use technologies such as the magneto-optic technique, which electrically alters the bias of grains of material after they have been heated by a laser. Compact discs (CDs) and laser (or video) discs are optical discs. Their storage capacities are far greater than magnetic media and are likely to replace magnetic hard disks and tape in the near future.

Optical disk—Medium that will accept and retain information in the form of marks in a recording layer, that can be read with an optical beam.

Optical scanner—Input device that translates human-readable or microform images to bit-mapped or raster machine-readable data.

Orientation—The relative direction of a display or printer page, either horizontal ("landscape" orientation) or vertical ("portrait" orientation).

Output device—Any device by which a computer transforms its information to the "outside world." In general, you can think of an output device as a machine that translates machine-readable data into human-readable information. Examples: printers, microform devices, video screens.

Overscan—The part of an image that falls outside the borders of the display screen, i.e., the part you cannot see.

Packet—A group of bits, package together, for transmission purposes. Three principal elements are included in the packet. (1) Control information, destination, origin, length of packet, etc.; (2) the data to be transmitted; and (3) error detection and correction bits. Sending data in packets rather than continuous streams offers more efficient use of transmission lines.

Page recognition—OCR software that can tell the difference between text on a page and other items, such as pictures, artwork, etc.

Pan—To view a different part of a page that has been overscanned (is off the borders of the screen).

Parallel—The transmission of bits over multiple wires at one time. Accomplished by devoting a wire for each bit of a byte. Parallel data transmission is very fast, but usually happens only over short distances (typically under 500 feet) because of the need for huge amounts of cable. Most often used in computer-to-printer, and scanner-to-computer applications.

Parallel processing—The simultaneous execution of two or more process operations in a single computer system.

Parallel routing—Parallel routing enables the route composer to send a route to many nodes at once. Each of the nodes can act individually on the information contained in the route. The workflow term for a parallel route is an AND-Split.

Parity—Used in error correction. A separate bit, the parity bit, is added and manipulated so that the number of 1s is odd (for odd parity) or even (for even parity). If the number of bits sent doesn't conform to the parity, the software detects an error. Parity is typically combined with data stored in positionally corresponding blocks of other disks in the RAID set to regenerate the missing data.

Parity bit—Redundant binary digit added to a series of such digits and assigned the value required to make the sum of all bits odd or even. See also parity.

Parity check—Test that helps ensure the validity of data by determining whether the number of zeros and ones represented by binary digits of a byte of data is odd or even.

Pattern matching—An OCR technique. With pattern matching, the software is given a "template" of possible characters. When the scanner sees a letter, it compares it to its library of pattern templates. If there is enough of a match, it safely assumes it has "recognized" the letter and sends the ASCII equivalent of the letter to the output file.

Persistence—A way to overcome flicker in a CRT that has a slow "refreshing" rate. The phosphors remain glowing, or "persist," after they have been energized.

Personal Computer (PC)—Used to indicate an IBM or compatible. Sometimes it is used more generally to indicate any personal computer.

Phase change recording—An optical recording technique. The laser strikes the medium and causes it to crystallize in a controlled way, thereby reflecting light to the reading laser.

Phosphor—Substance which glows when struck by electrons. The back of a cathode ray tube face is coated with phosphor.

Pitch—(1) The number of characters per inch measured horizontally. Fixed spacing printers have the same pitch for every letter, regardless of the letters' widths. Proportional spacing has varying pitch, depending on the letter. (2) The distance between grooves (measured center to center) on an optical disc.

Pixel—A sort-of acronym for " Picture Element". Also called a Pel. When an image is defined by many tiny dots, those dots are pixels. On the printed page, each pixel is one dot. On color monitors, though, a pixel can be made up of several dots, with the color of the pixel depending on which dots are illuminated, and how brightly.

Polymorphism—A programming technique used to enhance the reuseability of software components

PostScript—A software published by Adobe Systems that translates graphics created in a computer to language a (PostScript-compatible) printer can understand. It is called a page description language. Postscript-compatible printers have interpreters in them that create the proper dot patterns to recreate the screen image—text and graphics—to a page of paper.

Process execution—The duration in time when manual process and workflow process execution takes place in support of a process.

Process instance—Represents an instance of a process definition which includes the manual process and the automated workflow process.

Process role—A synergistic collection of workflow activities that can be assumed and performed by a workflow participant for the purpose of achieving process objectives.

Properties—Properties are the attributes associated with a component object.

Protocol—Formal set of conventions governing the orderly exchange of information between communicating devices by defining such things as connection establishment, security provision, data sequencing, error control, etc. Protocols achieve efficient line use by reducing the amount of information transferred by distinguishing between device control information and data

Proximity Search—A feature of full-text searching, in which every occurrence of a word within a certain distance of another word is found, i.e. finding every time the word "budget" is mentioned within 20 words of the word "Congress."

Pulse Code Modulation (PCM)—The most common method of encoding an analog signal into a digital bit stream. PCM refers to a technique of digitization, not a universally accepted standard of digitization.

Query—To ask or inquire about something within a database.

Queue—A stream of tasks waiting in line to be executed.

Queue time—the amount of time in which a case is waiting to be serviced, starting from the point that the task processing is started.

Radial acceleration—The rate at which a track on an optical disc accelerates toward and away from the center, because it is not perfectly aligned or perfectly round.

RAID rank—Group of drives managed by the disk array controller and configured to work as a defined set in host I/O operations.

Random Access Memory (RAM)—The primary memory in a computer. Memory that can be overwritten with new information. The "random access" part of its name comes from the fact that all information in RAM can be located, no matter where it is, in an equal amount of time. This means that access to and from RAM memory is extraordinarily fast. By contrast, other storage media, like magnetic tape, requires searching for the information, and therefore takes longer.

Raster—Description of a rectangular or square array formed by a number of horizontal lines comprising a number of picture elements. The number of scan lines establishes the vertical dimension of the array and the number of picture elements forms vertical rows which establish the horizontal dimension of the array.

Raster data—Set of data defining the values of pixels in a raster image.

Raster display—The most common type of display terminal. Uses pixels in a column-and-row array to display text and images.

Raster graphics—Method of representing a two-dimensional image by dividing it into a rectangular two-dimensional array of picture elements.

Raster image—Image formed by modulating the intensity of the individual picture elements within a raster array.

Raster line—Thin, horizontal strip across an image, captured one at a time by elements in the scanner.

Raster scan—Method of generating or recording the elements of an image via a line-by-line sweep.

Raster to vector conversion—Conversion of a raster image into a vector data image.

Readability—The degree to which an image on screen is clear and the content discernible to the average human eye at normal viewing distances.

Relational database—A database built and operated in accordance with the relational model of data which holds that all data be organized as a set of two dimensional arrays or tables which have a relation to each other.

Rewritable optical—Optical media from which data can be erased and new data added. Magneto-optical and phase change are the two main types of rewritable optical discs.

Read cache—The cache is used to accelerate read operations by retaining data which has been previously read, written, or erased, based on prediction that it will be reread.

Read Only Memory (ROM)—Data stored in a medium that allows it to be accessed but not erased or altered.

Record—In a database, a record is a group of related data items treated as one unit of information, for example, policyholder's name, address, social security number, etc. Each item in the record is a field.

Recording zone—The ring-shaped area of an optical disc on which data can be recorded.

Reduced Instruction Set Computing (RISC)—A computer system with a special microprocessor that processes fewer instructions, and thereby is much faster. A RISC system depends on software to perform many of the functions that would normally be done by microprocessors. RISC workstations are used in calculation-intensive operations such as those performed by computer-aided design (CAD) and computer-aided manufacture (CAM) engineers.

Redundant Arrays of Inexpensive or Independent Discs (RAID)—A storage device that uses several optical discs working in tandem to increase bandwidth output and to provide redundant backup.

Refresh—The phosphors at each pixel of a CRT which are stimulated by a charge from an electron gun glow only briefly. They must be renewed frequently in order for the image to appear stable. This renewal is called refreshing.

Refresh rate—Measure of how often the image on a CRT is redrawn; often expressed in hertz. Typically 60 times per second, or 60 Hertz (Hz), in the United States.

Relational database—A database built and operated in accordance with the relational model of data which holds that all data be organized as a set of two dimensional arrays or tables which have a relation to each other.

Remote Procedure Call (RPC)—A mechanism through which applications can invoke procedures and object methods remotely across a network. Using RPC, an application on one machine can call a routine or invoke a method belonging to an application running on another machine.

Rendezvous routing—Rendezvous routing, or joins, allows the consolidation of workflow route information. The two types of rendezvous routing are synchronous and asynchronous joins.

Resolution—(1) Measure of imager output capability, usually expressed in dots per inch (dpi). (2) Measure of halftone quality, usually expressed in lines per inch (lpi). The higher the resolution, the greater amount of detail may be shown. If a resolution is agreed upon as a standard, it is called a graphics standard.

Retirement—The term that describes the decision to throw away a recording medium (optical disc or mag tape) when it has too many defects.

Retrieval key—A word, number or phrase associated with a document to aid in its retrieval from storage. Sometimes called descriptors. There are often many retrieval keys used together to fully locate a document; together they are called an index.

Rewritable optical disk—Optical disk on which data is recorded. The data in specified areas can subsequently be deleted and other data can be recorded.

Rule—a definition criteria that the system will evaluate to automatically determine an action or route to be taken by a case at a particular point in a workflow process.

Scaling—Technique using an algorithm to convert a bitmap of one density into a bit map of another proportional density. Scaling usually involves enlarging or contracting an image.

Scan—To convert human-readable images into bit-mapped or ASCII machine-readable code.

Scan head—The part of the mechanism of a scanner that optically senses the text or graphic as it moves across a page.

Scan rate—Number, measured in times per second, a scanner samples an image.

Scan size—Dimensions (length and width) of the part of a document that can be digitized.

Scan time—Total time to convert text or graphical information to electronic raster form.

Scanner—(1) A device that optically senses a human-readable image, and contains software to convert the image to machine-readable code. (2) Device that electro-optically converts a document into binary (digital) code by detecting and measuring the intensity of light reflected or transmitted.

Scanner threshold—Setting that determines whether a pixel is white or black.

Scanning—(1) Operation which precedes digitization whereby the surface of a document is analyzed for characters and graphics, and analog signals are produced corresponding to the optical density of the sampled points. (2) OCR scanning is the conversion of printed or other symbolic information from paper or microform into ASCII code. (3) The systematic examination of data.

Screen—Series of dots (may also be series of lines or other pattern) used to represent continuous tone artwork.

Screen capture—To transfer what presently appears on a display screen to a computer file.

Scrolling—The image constantly rolling (moving up or down) on the display.

Sector—The smallest addressable unit of an optical disc's track. Contains 512 bytes.

Seek error—The inability of an optical drive to fine the user's request because of disc flaw or vibration or the drive just doesn't work right.

Self Coupled Optical Pickup (SCOOP)—An optical drive design that combines the functions of the laser reading device with the photodetector used to accept tracking and focus error signals.

Sensitive layer—The layer in an optical medium where the data is recorded; it may be composed of more than one layer or materials. It is sandwiched by protective and supporting layers.

Sensitivity—Measure of the light dose needed to mark an optical medium.

Sequential routing—Sequential routing enables the user to specify a workflow route in which the message goes from one route to another.

Serial—Data communications mode in which bits are sent in sequence.

Serial Storage Architecture (SSA)—A high speed serial interface designed and marketed by IBM.

Server—A computer which is dedicated to one task. A database or directory server would be responsible for responding to a user's search request, returning the list of stored documents that meets with the parameters of the request.

Servomechanism—Devices which constantly detect and adjust some variable. Optical drives have focus and tracking servers.

Shadow mask—A thin sheet of metal with tiny holes located inside a color monitor behind the phosphor. The three electron beams inside the monitor must shoot at the monitor's phosphor through a shadow mask to achieve color clarity or keep the three colored dots from overlapping or bleeding into each other.

Simplex—Method of recording images one by one in which a single frame appears within the usable width of the microfilm (ISO). See also image arrangement.

Skew—To slant a selected item in any direction; used in graphics and desktop publishing.

Slot—Refers to the number of available spaces in an optical disk jukebox for additional optical media to be stored. Also refers to the open interconnection inside a computing device where additional circuit boards could be attached.

Small Computer System Interface (SCSI)—An industry standard for connecting peripheral devices and their controllers to a microprocessor. The SCSI defines both hardware and software standards for com-

munication between a host computer and a peripheral. Computers and peripheral devices designed to meet SCSI specifications should work together. A single SCSI adapter card plugged into an internal IBM PS/2 micro channel PC slot can control as many as seven different hard disks, optical disks, tape drives and scanners, without siphoning power away from the computer's main processor. Formerly known as SASI (Shugart Associates Systems Interface).

Software—Set of programs, procedures and documentation concerned with the operation of a data processing system.

Source-document capture—Conversion of documents, usually paper, to microimages or digital images.

Speed—(1) Quantitative measure of the response of the sensitized material to radiant energy for the specified conditions of exposure, processing and measurement. (2) Maximum aperture of an objective lens. (3) Chemical activity of a processing solution.

Spin-up—The time during which a drive accelerates its disk/disc up to operating speed.

Spindle—The center part of a disk (or disc) drive which maintains the axis of rotation and provides the force to rotate the disk (or disc).

Stage—The predicted movement, behavior activity of images inside an image management system.

Storage capacity—Amount of data that can be contained in an information holding device or main memory, generally expressed in terms of bytes, characters or words.

Storage media—The physical device itself, onto which data is recorded. Mag tape, optical discs, floppy disks are all storage media.

String—A series of characters, usually the subject of a text search.

Stripe—The data and parity from the associated chunks of each member of the RAID set.

System—Organized collection of hardware, software, supplies, people, maintenance, training and policies to accomplish a set of specific functions.

Systems Network Architecture (SNA)—IBM's very successful means of networking remotely located computers. It is a tree-structured architecture, with a mainframe host computer acting as the network control center. Unlike the telephone network, which establish a physical path for each conversation, SNA establishes a logical path between network nodes, and it routes each message with addressing information contained in the protocol.

Tagged Image File Format (TIFF)—A bit map file format for describing and storing color and gray scale images.

Tape backup—Making mag tape copies of hard disk and optical disc files, for disaster recovery.

Terabyte—From "tera," which means trillion, although it actually means 1,099,511,627,776 bytes in a computer's binary system. A terabyte is 1,024 gigabytes.

Terminal—Any device, capable of sending or receiving information over a communications channel.

Text based—Representation of images that requires the use of pre-existing characters rather than vector or raster graphic techniques.

Text files—A data file consisting of alphanumeric characters, defined by a text format such as ASCII or EBDIC. Entries in a text file are available for text searching.

Text management—All the techniques and technologies involved in creating, storing and retrieving text files in an organized and logical manner.

Text search—A technique for examining text files for occurrences of specific sets of characters, either in a string (a word or sentence) or in proximity (a certain word in the vicinity of another word). A "contextual search" involves finding entire documents based on a string of characters that appears in it.

Text/image retrieval—The ability to locate a page image by using a full-text search.

Threshold—A predefined level set into a scanner's software to determine whether a pixel will be represented as black or white.

Thresholding—Process by which, in a photodetector, i.e photodiode, CD, etc., the analog gradation of dark to light is recognized by the scanner's detection mechanism to produce digital signals.

Throughput—The amount of time it takes for the processing of data from the beginning of a particular process, to the end of the process. Throughput also can refer to the number of items completed in the process.

Tiling—Reproducing oversized artwork or documents by breaking the image area into parts (called tiles). Adjacent tiles repeat a small portion of the image, and they may contain crop marks as well. The repeated portion of the image (the overlap) and the crop marks aid in reconstructing the overall image from the tiles.

Track—The path which is to be followed by the read header or beam during the magnetic or optical reading of a disk or disc; or the path to be followed by the recording head or beam during the writing of a disk or disc. In an optical system, the track consists of the Groove (recordable) and the Land (un-recordable).

Track Jump—The action of moving quickly from one track to another nearby.

Tracking server—The mechanism in an optical drive which senses and adjusts for variations in movement of the recording area (the groove) of a track, caused by imperfections in the medium of the drive mechanism.

Transmission Control Protocol/Internet Program (TCP/IP)—A set of protocols developed by the Department of Defense to link dissimilar computers across networks.

Tree-structured directories—A familiar name for hierarchical file management, used by both DOS and Macintosh operating systems. So-called because subdirectories can be thought of as "branching" away from the main, root directory.

Trichromatic—The technical name for RGB representation of color, i.e., using red, green and blue to create all of the colors in the spectrum.

Underscan—The part of an image that is inside the borders of the display screen, the part you can see.

Unfragmented—A hard disk that has most of its files stored in consecutive sectors, rather than spread out all over the disk in an interleaved fashion.

UNIX—A general-purpose, multi-user, multitasking operating system invented by AT&T. UNIX is powerful and complex, and needs a computer with a large amount of RAM memory to support its power. UNIX allows a computer to handle multiple users and multiple programs simultaneously. And it works on many different computers, which means you can often take applications software and move it, with little changing, to a bigger, different computer, or to a smaller, different computer. This process of moving programs to other computers is known as "porting."

Variable length record—A file in a database containing records not of uniform length and in which the distinctions between fields are made with commas, tabs or spaces (called "delimited.") Records become uniform in length either because they are uniform to start with or they are "padded" with special characters.

Vector—Images defined by sets of straight lines, defined by the locations of the end points. At larger magnifications, curves may appear jagged. This condition is called aliasing.

Vector data—Digital description of an image stored as a series of points and mathematical functions to describe the geometric figure, i.e., line, circle, arc, etc.

Vector display—Terminal that displays images with vectored line segments, rather than pixels.

Vector to raster conversion—Conversion of vector data image into a raster image.

Vectorization—Translation of a pixel-based image to a vector-based image, usually to be compatible with a CAD program.

Vertical recording—A magnetic disk recording technique that increases the available storage space.

Vertical scan frequency—Same as refresh rate, expressed in Hertz (Hz).

Very High Density (VHD)—Techniques of recording 20 megabytes and more on a 3 1/2" magnetic disk.

Video Graphics Array (VGA)—Standard IBM video display standard. Provides medium-resolution text and graphics.

Visual Basic Extensions (VBXs)—Add on programs (written in C or C++) to the Microsoft's Visual Basic development environment that perform specific functions, such as putting a menu on a screen, or deskewing images. VBXs can be combined to build customized document imaging application modules.

Volume label—A name assigned to a floppy or hard disk in MS-DOS. The name can be up to 11 characters in length. You assign a label when you format a disk or, at a later time, using the LABEL command.

Wand—A hand-held scanner used for OCR or for reading bar codes.

What You See Is What You Get (WYSIWYG)— Pronounced "wizzy-wig," it refers to a graphics or publishing program that displays images on the screen (nearly) exactly the way they will appear on paper.

Wide Area Network (WAN)—Communications network that links broad geographic areas.

Wildcard—A character in a text search that stands for other characters. For instance, a search for GEO* (with the asterisk being the wildcard) would find all occurrences of words starting with the letters GEO—geography, geostationary, geology, etc.

Windows—A Microsoft operating system that features multiple screens and a graphical user interface (GUI).

Workflow—A program that queues, tracks and otherwise manages documents, work items, and collections of documents and work items as they progress from entry into the system, through the various individuals or departments in the organization until a business process is completed.

Workflow application—A software program(s) that will either completely or partially support the processing of work items in order to accomplish the objective of a workflow process activity instance or instances.

Workflow enactment service—A software service that may consist of one or more workflow process engines in order to create, manage and execute workflow process instances. Client workflow applications/tools interface to this service via the workflow application programming interface (WAP).

Workflow management system—An electronic system that includes workflow capabilities to route, schedule and control business processes often triggered by the movement of documents within an organization.

Workflow participant—A resource which performs partial or in full the work represented by a workflow process activity instance.

Workflow process—The computerized facilitation or automated component of a process.

Workflow process activity instance—An instance of a workflow process activity that is defined as part of a workflow process instance.

Workflow process control data—Data that is managed by the Workflow Management System and/or a Workflow engine.

Workflow process definition—The component of a process definition that can be automated using a workflow management system.

Workflow process engine—A software service or "engine" that provides part or all of the run time execution environment for a workflow process instance.

Workflow process execution—The duration in time when a workflow process instance is created and managed by a Workflow Management System based on a workflow process definition.

Workflow process monitoring—The ability to track workflow process events during workflow process execution.

Workflow system—A system that automates the processing of documents, scheduling, processing, routing documents automatically among departments and tracking document status.

Workflow templates—Workflow templates are pre-programmed route-enabled messages that are used to create new instances of routing messages.

Work item—Representation of work to be processed in the context of a workflow process activity in a workflow process instance.

Work item pool—A space that represents all accessible work items.

Worklist—A list of work items retrieved from a workflow management system.

Worklist handler—A software component that manages and formulates a request to the workflow management system in order to obtain a list of work items.

Work management system—If rules concurring task triggers, task durations, and intertask dependencies are combined with monitoring agents, a work management system can be created. The work management can generate triggering requests when tasks complete. It can generate reminding, reporting, and resolving tasks when tasks are slipping. And it possesses sufficient information to provides up-to-the-minute status reporting on demand.

Work process—Any process that involves cooperative work; that is, processes involving multiple persons working to accomplish a specific goal.

Workstation—A single-user microcomputer or terminal.

Work task—A work process may be broken down into work tasks: units of coordinated work between exactly two roles. Any work process can be conceptualized as a group of interacting work tasks.

Write back cache—A cache write strategy that writes to the cache memory, then may flush the data to the primary media at some future time. The user sees the operation as complete when the data has reached the cache. The intent of this strategy is to avoid unnecessary accesses to the primary media.

Write Once Read Many (WORM)—Optical storage device on which data is permanently recorded. Data can be erased, but not altered, and no additional data can be added.

Zoom—To enlarge a portion of an image in order to see it more clearly or make it easier to alter.

This Technology Guide is one of a series of guides, written by The Rheinner Group and published by ATG, designed to put complex document management, imaging, and workflow concepts into practical and understandable terms. Each guide provides objective, non-biased information to assist in the internal education, evaluation and decision making process. This Technology Guide, as well as the other Document Management, Imaging, and Workflow Technology Guides in the series, are available on ATG's Web Site.

www.techguide.com

A joint venture

ATG
The Applied Technologies Group

The Rheinner Group

Tel: (508) 651-1155
Fax: (508) 651-1171
E-mail: info@techguide.com

Tel: (617) 741-8100
Fax: (617) 741-5885
E-mail: info@rheinner.com