# NASA

Scientific and Technical
Information Facility

# DIGITAL DOCUMENT IMAGING SYSTEMS:

# AN OVERVIEW AND GUIDE

AUGUST 1, 1990

operated by:

**RMS ASSOCIATES**

NASA

Scientific and Technical Information Facility

# DIGITAL DOCUMENT IMAGING SYSTEMS:

# AN OVERVIEW AND GUIDE

AUGUST 1, 1990

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

This document was written for the NASA Automated Information Management (AIM) Program Office in cooperation with the Headquarters Resources Management Council (ARMC) to assist managers in objectively understanding the technology and pitfalls that are part of the current environment of digital document imaging.

We have identified the key components of a complete Digital Document Imaging System (DDIS) and noted the areas that require careful attention when selecting such a system. (See *Appendix A* on page 33 for more detailed information.) Because the technology is new and each vendor seems to have a "better mousetrap," it is extremely important to define requirements first and then survey the marketplace to determine if the technology satisfies them. In addition, because the technology is new and ever changing, it is important to be aware that very few standards allow for the interchangeability of information between various systems. For example, it might not be possible to exchange documents from one system to another because of system and data-format incompatibilities.

With this document we hope, through the ARMC, to establish guidelines to help insure compatibility amongst the various systems where ever it is possible.

We would like to express our thanks to the following employees of RMS Associates for their efforts in producing this document: Jean M. Tolzman, Ralph E. Daniel, Gregory C. Decker, and Bruce K. Ansley.

If there are any questions or comments concerning this document, please contact me at 202/453-1790.

Pete Messina
NASA Headquarters Code NTI
August 1, 1990

# 1. INTRODUCTION AND OVERVIEW

## Introduction

The purpose of this document is to aid NASA managers in planning the selection of a Digital Document Imaging System (DDIS) as a possible solution for document information processing and storage. Intended to serve as a managers's guide, this document contains basic information on digital imaging systems, technology, equipment standards, issues of interoperability and interconnectivity, and issues related to selecting appropriate imaging equipment based upon well-defined needs.

## Overview

From a user's perspective, the development of a Digital Document Imaging System appears relatively straightforward. Simply take a computer, add an input device (such as a scanner), a storage device (such as an optical disk drive), and an output device (such as a printer), and put them all together. Each component in this system provides a specific function. Figure 1 at right shows the relationships of these functions. Putting these individual components together to create a Digital Document Imaging System encompasses a complex interrelationship of hardware, software, industry standards, and user requirements.

**Figure 1.** *DDIS Primary Application Functionalities*



Every day organizations create documents, such as manuals, training information, litigation support, compliance reports, and inquiries as well as letters, memoranda, contracts, brochures, invoices, receipts, data, text, photos, drawings, and handwritten notes. These documents are then word processed, photocopied, mailed, facsimile (fax) transmitted, filed, stored, and retrieved. Mostly in paper form, these documents usually require immense storage space and repeated physical handling. Many organizations that want to improve handling, storage, and archiving of documents use Digital Document Imaging Systems. Although seductive as a cutting-edge technology and promising as a tool for document automation, Digital Document Imaging Systems might not always be the cure-all for automating some paper-intensive processes.

The vision within many Information Resources Management (IRM) divisions is to obtain Digital Document Imaging Systems that can provide organization-wide document exchangeability, system interoperability, and, where possible, cooperative processing, shared files, shared resources, and reduced labor and operating costs. For this IRM vision to be realized, not only must a system be chosen carefully, but also close attention must be paid to the types of documents processed and the processing performed.

## 1.1. SCOPE

This guide, which is designed to present essential information to have when considering a Digital Document Imaging System, is divided into the following five sections:

Section 1, *Introduction and Overview*, is this section, which describes the content of this guide and contains brief definitions of key terms.

---
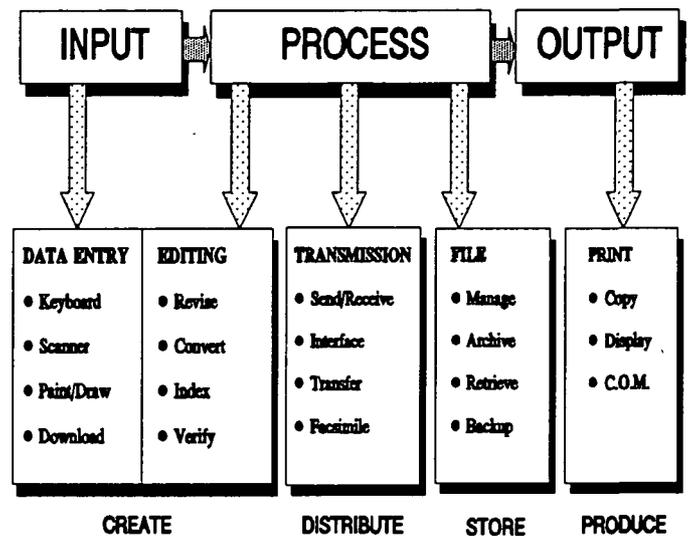
Section 2, *Considerations*, provides basic information to help managers decide if Digital Document Imaging Systems are appropriate solutions to address their needs. This section also guides managers towards Digital Document Imaging Systems that will comply with long-term objectives regarding interoperability and interconnectivity for equipment and data.

Section 3, *Equipment, Formats, and Standards*, contains a summary of DDIS hardware and software components, data formats, and the standards that govern both equipment and data formats. Issues include standards compliance, document portability, equipment interoperability, and future directions in standards.

Section 4, *Technology Summary and General Guidelines*, provides guidelines for NASA managers to use when planning a Digital Document Imaging System. Included are descriptions of analytic tools and processes that will help clarify functional and systems requirements. In addition, general recommendations are provided for both short- and long-term strategic systems planning.

Section 5, *Appendices*, provides details on Digital Document Imaging System component specifications (*Appendix A*), organizations that issue applicable standards (*Appendix B*), cross-references to these standards (*Appendix C*), additional resources to consult for further information (*Appendix D*), and the terminology in use (*Appendix E*).

## 1.2. DEFINITION OF KEY TERMS

To be able to discuss important concepts of Digital Document Imaging Systems, it is necessary to understand the meaning of terminology used in digital imaging. Key terms in this rapidly evolving field are defined below. *Appendix E* on page 69 contains an expanded glossary.

**Compatibility:** The characteristic of computer equipment and software by which they are executable or translatable on more than one class of computer without data conversion or code modification. Compatibility can include electrical, logical, operational, and mechanical characteristics as applied to computer programs, system components, input/output (I/O) devices, and interface equipment.

**Connectivity:** In communications, the ability to make a network connection that allows integrated network products to work in the same environment through standards and protocols, such as Open Systems Interconnection (OSI), Integrated Services Digital Network (ISDN), and X.25. Connectivity makes interoperability possible.

**Cooperative Processing:** The ability of two or more programs to interact dynamically to accomplish a common processing objective that may involve heterogenous processing environments. A Client/server application is an example of cooperative processing; connectivity makes it possible.

**Digital Document:** An external document that has been converted to a computerized digital data representation. This representation may be in a coded text format if the document contains text or in an image format.

**Digital Image:** The bit-mapped representation of a document as an image. Digital images are created by using either an image editing program or digital scanner.

**Exchangeability:** The ability to interchange text and images among multiple computer hardware and software platforms.

**File Transfer:** The software process of sending a data file from one computer system to another.

**Image Format:** The encoding scheme for a bit-mapped raster image.

**Interface:** A common interconnection between systems, components or functions that normally do not interact.

**Internetworking:** The ability to communicate across different networks between end systems.

**Interoperability:** The ability to initiate and control processing across different systems without any discernable operational differences.

**Open System:** The ability of an end-system of one design to connect with any other end-system conforming to the reference model of International Organization for Standardization (ISO) Open Systems Interconnection (OSI) and the associated standard protocols.

**Portability:** The ability of applications to execute on different systems without data conversion or code modifications. Necessarily involved are standards governing operating systems, database management system (DBMS), data exchangeability, network services, user interfaces and programming services.

**Protocol:** In communications, a protocol is a rule or management method for the data link and its associated attached terminal devices. Protocols permit the exchange of information by establishing a series of rules for interpreting control characters.

**Standards:** Defined rules accepted and followed by a group having similar needs that allow a common method of exchange to take place. Standards can be subdivided into voluntary and regulatory standards. Voluntary standards are market driven and unenforceable; regulatory standards are usually more rigid, more clearly defined, and more enforceable.

# 2.  CONSIDERATIONS

A Digital Document Imaging System (DDIS) can be used to address the problem of handling and storing large volumes of documents. Although the concept of a Digital Document Imaging System is not new, until recent advances in data processing and automation, these systems were generally too large and expensive to consider as document automation solutions.

The impression often given by sales representatives and their glossy brochures is that a Digital Document Imaging System can solve any document automation need. The unsuspecting buyer, who often wants a quick, easy solution, can be left with a system that will never fulfill expectations, much less actual needs. This section provides guidelines to help managers determine if a Digital Document Imaging System is an appropriate document automation solution that will meet their needs. This section also contains information to consider when planning the acquisition of a Digital Document Imaging System so it can meet important long-term objectives regarding data exchangeability and system interoperability.

## 2.1.  WHEN TO CONSIDER A DIGITAL DOCUMENT IMAGING SYSTEM

Three major technological factors have contributed to the rapid rise of digital imaging: computer processing power, improvements in communication, and reduced data storage costs. The computers available today are smaller, less expensive, and have greater ability and capacity than their predecessors of 10 to 30 years ago. Concurrent with computer enhancement, communication capabilities have greatly improved while the cost for these services on a per-unit basis has dropped. At the same time, storage media costs have declined while the capacity, flexibility, availability, and the variety of media have significantly increased. When combined, these factors make Digital Document Imaging Systems attractive solutions to document automation problems. Without good analysis, planning, and understanding of the limitations of this new and enhanced technology, the attempted solution could prolong, if not worsen, the problem.

To determine whether or not a Digital Document Imaging System is a viable document automation solution, each circumstance must be considered individually. Some situations, such as heavy reliance on changeable, word processed text may be best served by a traditional computer-based word processing system. For fixed format, infrequently modified documents that are frequently viewed online and printed, a Digital Document Imaging System can offer near-instant access to documents. These systems are effective means of storing graphics, but, short of cutting and pasting images, they currently offer limited editing capability. Optical storage devices (such as optical disks and CD-ROM), which are not limited to use in Digital Document Imaging Systems, offer extended storage life, higher-density storage, greater ease of archiving, and lower costs per unit of storage. However, optical devices are slower than magnetic-media devices (such as hard-disk drives) and can require special systems integration hardware and software.

An organization's application could possibly benefit from a Digital Document Imaging System if the organization has any of the following needs:

- To process incoming documents faster
- To process a large volume of documents
- To route incoming documents to various employees, or to require scheduling, distributing, and monitoring of these documents
- To process documents that contain nonchanging image information (signatures, pictures, or notes)

PRECEDING PAGE BLANK NOT FILMED.

- To access stored documents quickly (for example, customer response)
- To support simultaneous multiple access to a document
- To prevent serious consequences that would result if a document were lost, missing, or misfiled
- To print copies of documents on demand
- To make best use of limited or expensive or physical storage space
- To integrate document storage/retrieval with a transaction-processing system

Merely scanning a document into a Digital Document Imaging System without additional processing allows the document to be handled as an image only—it does not allow the document to be manipulated by traditional word processing. These images, or digital photographs, are bit-mapped images of full pages. These bit-mapped images require increased storage and transmission space when compared with the same document produced by a word processor. Also, these images might not be convertible back to a character document without rekeying or labor-intensive processing. These images also are not machine readable for searching and indexing purposes. However, with a thorough understanding of needs, planning, analysis, and appropriate NASA guidelines, both data exchangeability and system interoperability are possible.

A Text Information Management System (TIMS) offers enhanced search and retrieval using document scanners, full-text OCR, databases, and sophisticated indexing techniques. A fully implemented TIMS allows extensive keyword searching, including Boolean logic and proximity searches. Proximity searching logically seeks or limits the terms to search within the same document (for example, FIND "Digital Document" ADJACENT TO "Imaging Systems"). Without a TIMS, imaging systems cannot perform content-based searches and thus simply cannot find all stored documents that contain a specific word or phrase.

Digital Document Imaging Systems can offer the following advantages:

- Reduced physical storage (such as filing cabinets)
- Better document standardization, control, and filing
- Faster document retrieval
- Improved document sharing and distribution
- Reduced paper, labor, and storage costs
- Reduction of duplicate documents
- Improved workflow
- Improved access to information

Digital Document Imaging Systems, which are generally limited in text- and graphic-editing capability, require careful planning and implementation. In addition, a Digital Document Imaging System and possible office retooling, retraining, and process reorientation.

## 2.2.  OBJECTIVES FOR DIGITAL DOCUMENT STORAGE SYSTEMS

When considering data-processing in a Digital Document Imaging System, an important long-term objective from an Information Resources Management (IRM) perspective is to facilitate an organization-wide information system that offers an environment of uniform computing and communications. An equally important near-term objective is to ensure that equipment and systems purchased and implemented today

provide future value and remain compatible with later system procurements. Data exchangeability and system interoperability provide benefits beyond cost savings—they are both essential to future compatibility and value.

Within any large organization, documents, which consist of text, charts, graphs, drawings, and so on, are exchanged among departments for comment, information, approval, filing, and distribution. Currently, many of these documents are stored on paper, but they could be relegated to a Digital Document Imaging System as long as the document exchangeability remains unhindered.

Consider the daily movement of Government documents between departments. Such documents could be composed of text, pictures, line art, and tables. If the Digital Document Imaging System of one department is not compatible with that of another department, exchange would have to be made on paper or might involve a complex conversion process that conceivably could alter a document's format or content, compromising the document's integrity.

True system-to-system compatibility would allow a direct exchange of documents, using magnetic or optical disks as the exchange media, without being concerned about hardware or file-structure differences. A higher level of compatibility would allow document exchange over a network; thus, documents on one system can be transferred electronically to another system. An even higher level of compatibility would allow complete interoperability between systems. At this level, the users of one system have access to data and software on another system, without having to exchange files first.

At the highest level of compatibility and interoperability, four combinations of local and remote operational processing (software execution) and data residency equally coexist. Thus, it becomes immaterial where the physical software and data are located, thereby attaining a seamless integration of hardware, software, and data. This level of integration and operation is essential for an organization-wide information system to succeed with maximum benefit to its users. This level can only be reached if Digital Document Imaging Systems are carefully chosen, based upon need, analysis, standards, and long-range objectives regarding data exchangeability, system interoperability, and preparation for an organization-wide information system.

# 3. EQUIPMENT, FORMATS, AND STANDARDS

A Digital Document Imaging System is an integration of hardware (used to input, store, and output data), software (used to process data), data formats (used to structure data for processing), and standards (used to ensure internal and external compatibilities). This section supplies equipment requirements, examines data formats and data exchange, and gives an overview of standards applicable to Digital Document Imaging Systems.

## 3.1. EQUIPMENT

Usually, digital image data cannot be added to an existing information system by simply declaring a new data type. Data resulting from digital imaging requires special enhancements that extend the basic computer system need to include high-resolution monitors, additional memory, and special software.

Digital Document Imaging Systems can range from basic to complex, and range from small to large. Basic imaging systems store and retrieve documents—functioning as electronic filing cabinets. Complex imaging systems provide document indexing, modification, and enhancement—functioning as highly integrated, organization-wide systems.

A Digital Document Imaging System usually consists of the following components:

- Devices for scanning, viewing, printing, and storing data
- Computer software
- One or more general-purpose computers

A sample Digital Document Imaging System is shown in Figure 2 on page 12.

### 3.1.1. Document Capture

External documents—that is, paper- or microform-based documents—are digitally stored internally in a computer as coded or uncoded information (or sometimes both). Traditional computer information—data or text—is considered to be coded information, represented at the symbol level by a finite character set (for example, the 8-bit ASCII code). A pattern not defined in the set cannot be represented. Uncoded information is merely a bit-mapped image that includes nonalphanumeric information, such as graphs, drawings, line art, illustrated, and handwritten documents, that is not represented by a symbolic pattern. Optical Character Recognition (OCR) is the automated process of identifying a portion of a bit-mapped image as a coded pattern representing an alphanumeric character. These coded patterns are combined to form a file of alphanumeric character that is recognizable to the human eye as text.

A Digital Document Imaging System needs a method to convert documents into bit-mapped digital images. This function is usually accomplished with a bilevel or grayscale scanner. Either paper or microfilm media can be scanned to create digital images. Several computer-assisted methods, such as a video digitizer board or bit-mapped paint program, can be used to input documents. Facsimile (fax) devices can also be used to capture document images electronically.

**Figure 2.** *Sample Digital Document Imaging System Configuration*
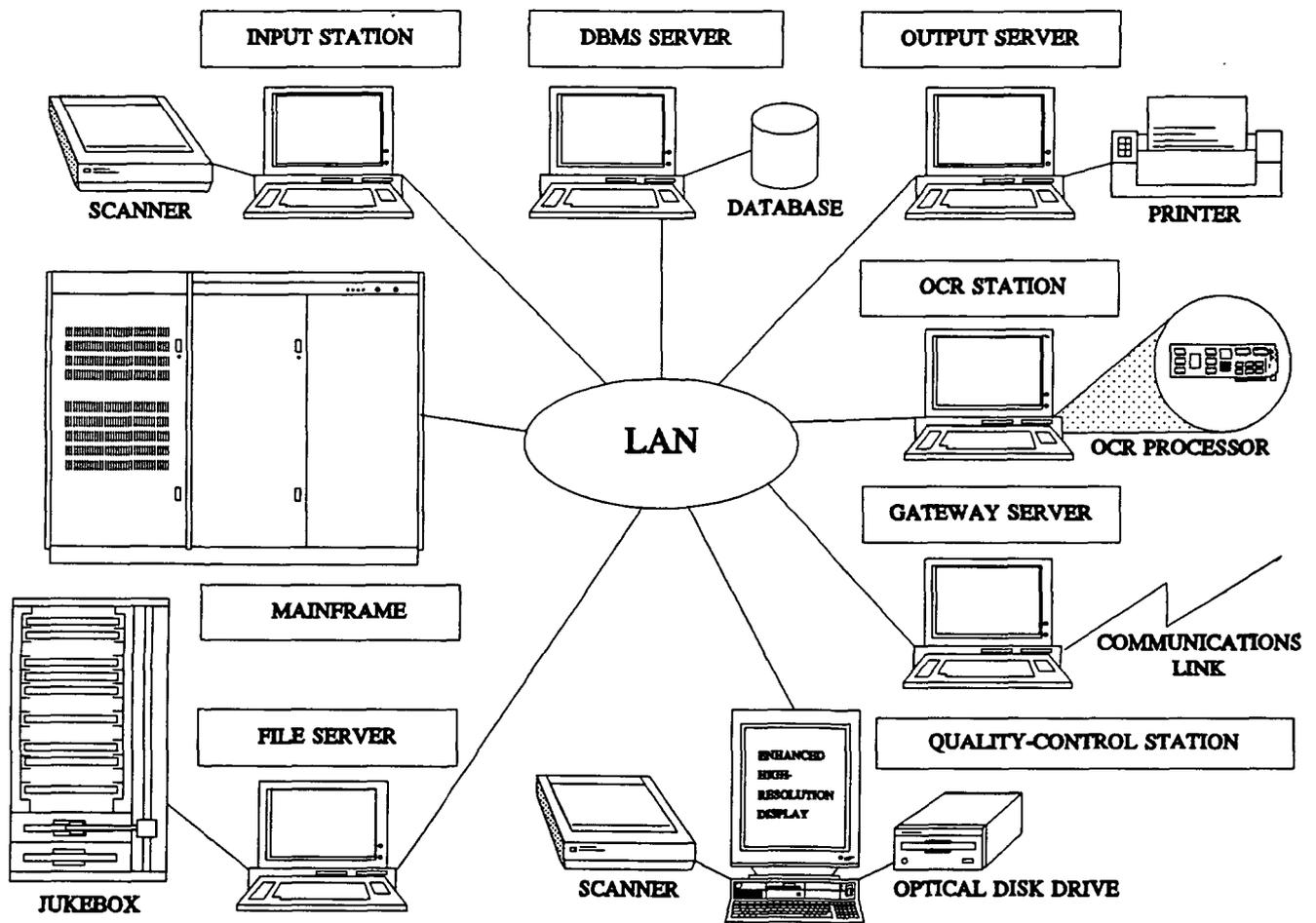


Additional input processes, including indexing and monitoring the quality of digital images, can be performed to enhance retrieval and improve the quality of stored images. Once digitized by scanners or fax machines, images can be converted to conventional ASCII character strings using an OCR process applied with an input processing system.

The quality of the internally stored document image is directly proportional to the quality of the original external document. If the original document is of poor quality, it could be difficult to apply OCR effectively. It is possible to use image-enhancement software to improve the quality of the scanned document image, but this type of software is no substitute for a high-quality original document image.

### 3.1.2. Document Holding

Documents stored as digitized images require considerably more space than documents stored using conventional coding, such as ASCII text or data. The average 2,000-character textual page stored conventionally requires about 2,000 bytes. The same document page, when digitally entered via a scanner and compressed for storage, requires 25 times more space, or about 50,000 bytes.

Magnetic, random-access, hard-disk storage is preferred for processing because of its fast access speed. However, optical storage can be better suited to hold image data because it is frequently less expensive per megabyte than magnetic storage. A 5.25-inch optical disk can store as much as several thousand pages of digitized images. WORM and CD-ROM disk storage are preferred for archival purposes because of their relative permanence (5 to 100 years). Optical-disk manufacturers have performed rigorous aging tests on these media to demonstrate their longevity, but only with the actual passage of time can proof of permanence be positively established.

CD-ROM improves upon the error detection and correction of audio-based CD players by adding an additional layer of Error Correction Code (ECC), which brings the error rate down to 1 in $10^{12}$—well within the tolerance demanded for use in digital storage. Even so, a working committee of the U.S. Justice Department investigating imaging technology has yet to rule on the acceptability of documents stored on CD-ROM as possible legal evidence because, in theory, the information in an original external document could be deliberately altered during the multistep premastering process before it is stored on CD-ROM. The ability of CD-ROM to keep information accurately stored is not in question. This same committee has indicated that WORM probably will be acceptable as legal evidence.

Jukeboxes allow an optical disk drive to access several optical platters. A jukebox provides greater flexibility, increased online storage, and reduced cost when compared with the equivalent storage capacity of other optical devices.

Storing information as text or data instead of an image offers real benefits, including the ability to alter and search this information. Text and data do not require the special hardware and large storage space that images do.

Sometimes, it is necessary to retain the original image in addition to the text or data. Many systems can handle compound documents—a combination of text and images. Some database management systems (DBMSs) include image data types. It is possible to store images and then run them through an OCR device later, but this capability must be engineered into the system.

### 3.1.3.  Document Indexing and Retrieval

Textual documents stored in a Digital Document Imaging System are of little value if they cannot be easily identified, searched, located, and retrieved. Digital Document Imaging Systems use databases to store documents and database indices to identify, locate, and categorize the documents within the database. Methods for data storage and organization include keyword systems, variable-length databases, document image processors, and full-text retrieval systems.

Keyword systems build indices, using predefined terms, of records containing these keywords. These systems limit searching to index terms only.

Variable-length databases provide for fixed- and variable-length test data fields, but are inefficient with large quantities of data. In addition, their search capability is limited to the indexed fields.

Document image processors input documents via optical scanners, frequently rely on a manually input (keyed) subject description data, and are limited to searches of the subject index.

Full-text retrieval systems consist of document image processors that employ Optical Character Recognition (OCR) software with complete indexing. Because all text is available, a database-oriented, full-text retrieval system provides the best scheme for document searching, storing, locating, and retrieving. Full-text retrieval systems allow keyword and context searching and can have thesauri to aid searching.

### 3.1.4. Document Presentation

A Digital Document Imaging System offers a capability for viewing a digital image that reflects the original or stored document page. The following devices are critical to this capability.

#### 3.1.4.1. Displays

With few exceptions, a Digital Document Imaging System handles black-and-white images and thus requires only a monochrome monitor. PC or workstation monitors capable of displaying graphics images, usually have adequate resolution for users to view document images clearly. Dual-page monitors are available that can show not only an entire 8-by-11.5-inch single page, but also two pages side-by-side. Monitors using Video Graphics Array (VGA) achieve a resolution of 640-by-480 pixels, or approximately 75 dots per inch (dpi). Although this resolution is satisfactory for viewing text, critical image analysis requires the use of a higher resolution display. Special-purpose monitors with a resolution of 150 dpi are available for image-quality verification, such as that needed for photographs, fingerprints, or signatures. A special-purpose graphics card and software, which normally includes routines for image viewing and manipulation, are usually bundled with these monitors to display images at the higher resolution.

#### 3.1.4.2. Printers

Just as optical disks make mass storage of digitized images possible, laser printers make fast, affordable, and detailed reproduction from digitized images practical. Digital images of photographs, charts, and drawings are usually printed at a resolution 300 dpi or higher for increased clarity. Text-only documents can be printed at lower resolutions between 75 and 150 dpi, providing adequate legibility and increased printer throughput. A critical element for selecting a printer for production use is its rated duty cycle—an indicator of the ability of a printer to handle reliably the volume of work required.

### 3.1.5. Processing

Digitized images do not require special processor hardware, but a file containing an uncompressed digitized image of a document is about 500 times larger than a file that has the same document stored only as a conventionally coded representation (such as ASCII) of the text. This large file size requires more powerful processors to maintain acceptable performance levels.

Two types of processing activities in a Digital Document Imaging System can have an impact on the speed with which digital images are manipulated—the central processing unit (CPU) and the operating system (OS). The CPU is the brain of a computer—all activity is controlled by this device. The operating system is an instruction set through which an individual communicates with a computer and the instructions that a CPU uses to communicate with its peripherals.

#### 3.1.5.1. Central Processing Unit (CPU)

The enormous size of digitized images requires the utmost in processor speed. The performance of a CPU that can acceptably process text can be severely degraded if it is tasked to process image data. Enhanced image processing hardware is a part of the usual Digital Document Imaging System. This addi-

tional and/or augmented hardware not only diverts work from the main CPU, but usually includes a built-in compression/decompression capacity to reduce the size of digital images, improving transfer rates, and diminishing storage requirements.

### 3.1.5.2. Operating System (OS)

As with hardware, digitized images usually do not require special operating systems or features, although some DDIS devices require special device drivers (software) or operating system extensions. The complexity and sophistication of most imaging applications, particularly when part of an integrated workflow, suggest that a multitasking operating system be used to optimize system flexibility and throughput.

### 3.1.6.  Typical Digital Document Imaging System Architectures

A Digital Document Imaging System can be classified on one of the following four levels:

- a.   Stand-alone storage and retrieval system
- b.   Image server to a mainframe system
- c.   Client/server integrated image system
- d.   Integrated image system

### 3.1.6.1. Stand-alone storage and retrieval system

The stand-alone storage and retrieval system is an electronic filing cabinet for image information. Its aim is for the direct storage and retrieval of images that have no integrated data or text relationships as part of the system. The system provides for the indexing of image information and usually reflects an existing archival, record management, or library application. This type of system is not suitable for transaction processing, which requires links to existing data and text services.

### 3.1.6.2. Image server to a mainframe system

In this type of system, the image server attaches to an existing time-sharing minicomputer or mainframe data-processing host system. Images and data retrieved from the host can be presented on a workstation monitor, usually in separate windows. Although this level of implementation sometimes can be awkward or complicated, it can be cost effective. With an image server, transaction processing is also possible and image processing can include OCR.

### 3.1.6.3. Client/server integrated image system

This level is an integrated environment that presents the user with one access and one window. A mainframe or minicomputer is usually involved but searches made within the imaging application will retrieve both image and host data transparently to the user. These systems frequently are communications intensive and are commonly used for claims-processing and customer-service applications. The image processor is capable of passing modified information back to the host for providing automatic mainframe record updating. This system is well suited for transaction processing.

### 3.1.6.4. Integrated image system

An integrated image system often is a complete integration of images with other components of text, data, and voice. Full access for information needs of electronic mail, software applications, digital and

analog images, and voice are available. Whatever the inquiry, complete information is quickly accessed and can be sent anywhere in any form desired or required.

## 3.2.  FORMATS

It is the coded, standardized format of the data—a digital document—that makes a compatible conversion-free data exchange possible. Unlike database and spreadsheet data, digital image data have not had the benefit of dominant industry leaders who have propelled *de facto* standards, such as the database file format standard of Ashton-Tate's dBASE or the spreadsheet file format of Lotus 1-2-3. The result has been the emergence of many data file formats, each more or less proprietary, to fill a particular software vendor's needs.

Data formats can be divided into the following four groups:

- Text file formats
- Raster (bit-mapped) graphics file formats
- Vector (object-oriented) graphics file formats
- Page Description Languages (PDLs)

### 3.2.1.  Text

Within a computer, the structuring of data is performed according to the architecture of the machine. This internal representation is not necessarily suitable for transmission or exchange outside of the machine; therefore, suitable codes were designed for this purpose. These codes include Baudot (a 5-bit code to replace Morse), Binary Coded Decimal (BCD, a 6-bit code), American Standard Code for Information Interchange (ASCII, a 7- or 8-bit code), and Extended Binary Coded Decimal Interchange Code (EBCDIC, an 8-bit code). In data transfers, avoiding needless code conversions usually improves efficiency and accuracy.

The predominant text format is ASCII, an standard adopted by the American National Standards Institute (ANSI). IBM extended the original ASCII code of 128 characters to 256 when it introduced the IBM PC in 1981. This extended set called the IBM Extended Character Set or the Extended ASCII Character Set.

Regarding standards, graphics formats are a microcosm of text formats—an evolution of market-driven formats created for specific purposes. To improve exchangeability and support open systems, graphics formats are now just beginning to gain the attention of standardization organizations and committees.

### 3.2.2.  Raster Graphics

Raster formats, also called pixel formats, are bit-maps of the dots that make up a graphics image. Each bit represents the on/off (binary) state of each pixel or dot. This bit-mapped representation, when used with images, is the derivative of the term digital. Scanners and paint programs directly create bit-mapped images, whereas photographic and video (analog) images can be converted to digital images by using a digitizing device or an autotracing program.

Bilevel or bitonal images, such as line art, contain only pure black and white—they have no grayscale values (stepwise variations from black to white). Scanners create bilevel bitmaps using thresholding—the process of selecting a cutoff level above which a pixel will be made black and below which a pixel will be made white. The stored bilevel image thus contains one bit per pixel. Scanners can have dynamic thresholding that can adapt the threshold automatically to produce the optimal contrast between the light and dark areas of an image.

Because the pixels on a computer display screen and the dots on a laser printer are typically of a fixed size, they cannot be used to produce a true halftone image, which creates in print (such as the pictures in a newspaper) the illusion of a continuous-tone image by using varying sized dots that correspond to the shades in the image. One way to represent a continuous-tone image with fixed-size dots is to assign grayscale values to the image. Scanners can capture continuous-tone images as grayscale values by using multiple bits to represent each pixel. A 4-bit per pixel scanner can capture 16 grayscale levels; an 8-bit per pixel scanner can capture 256 grayscale levels. (The human eye can distinguish between 50 and 100 grayscale levels in an image.) Grayscale levels are useful in synthesizing color images on monochrome displays and printers.

Tagged Image File Format (TIFF), originally developed by Aldus Corporation, is a bit-mapped format as well as a file format. It is widely use by scanner manufacturers and can be considered a nonproprietary *de facto* industry standard. TIFF images are saved in the same way as pixel-based programs, but can also be stored as grayscale values and contain information necessary for variable output resolutions.

Other common raster formats include the following:

- **CUT**       Media Cybernetics Dr. Halo format
- **GEM**       Digital Research Graphics Environment Manager format
- **MAC**       Apple Macintosh Macpaint format
- **MSP**       Microsoft Windows Paint format
- **PIC, PCX**  Zsoft PC Paintbrush format

### 3.2.3.   Vector Graphics

Vector formats, also called object-oriented images, contain the starting and ending points—the vectors—of each shape (points, lines, and polygons). A vector file is a set of instructions to construct an image rather than a representation of that image. Vector formats are used extensively in computer-aided design (CAD) systems because they efficiently store line drawings and are required by most plotters. Digital documents usually are not stored in a vector format.

Common vector formats include the following:

- **CGM**   Computer Graphics Metafile (Federal Information Processing Standard 128)
- **DIF**   Autodesk AutoCAD Drawing Interchange Format
- **DXF**   Autodesk AutoCAD Drawing Exchange Format
- **DRW**   Lotus Freelance file format
- **HPGL**  Hewlett-Packard Graphics Language (for HP plotters)
- **PIC**   Lotus 1-2-3 graphics (and others)

- **PICT**    Apple Macintosh picture file format
- **WPG**    WordPerfect Graphics file format
- **WPM**    Microsoft Windows Metafile

### 3.2.4.    Page Description Languages (PDLs)

The popularity of laser printers has led to the use of Page Description Languages (PDLs). These formats create files that are both bit-mapped and vector-based (for example, TIFF or Windows Metafiles). Although widely supported, the vector-based Encapsulated PostScript (EPS) format is generally considered to be inappropriate for data exchange because it produces very large and complex files.

Common PDLs include the following:

- **EPS**            Adobe Systems Encapsulated PostScript
- **DDL**            Imagen Document Description Language
- **PCL**            Hewlett-Packard Printer Command Language
- **TrueImage**    Apple/Microsoft image format (formerly called Royal)

## 3.3.    STANDARDS

Standards can avoid or minimize problems, much like a common language can enable a culture communicate more effectively. Standards help provide continuity to protect systems from the erratic conditions of the open market and enable equipment from different manufacturers to work together without previous collaboration. The benefits of standards compliance in Digital Document Imaging Systems operate over many tiers and provide the basis for exchanging electronic image media, interconnecting image equipment, sharing peripheral devices, and interoperating among different remote systems.

The U.S. Government is the largest purchaser of imaging systems, and its purchases guide the development and institution of applicable standards. In addition to the Government's purchasing power, policies such as the U.S Department of Defense (DoD) Computer-aided Acquisition and Logistics Support (CALS) initiative will likely promote the institution of standards applicable to Digital Document Imaging Systems. Although not yet fully implemented, CALS is nevertheless a significant step towards formal, established standards for information exchange.

An increasingly open world economy also could influence the extent to which U.S. manufacturers pursue standards. The 1992 European Standards program requires that Common Market Europe harmonize its standards. Many U.S. companies that sell products in Europe will have to abide by European standards to continue to sell in that market.

At the system level, Digital Document Imaging Systems do not have existing standards. At the system component and subcomponent level, the standards that exist, a majority of which are industry *de facto* standards, continue to evolve. The Association for Information and Image Management (AIIM) currently is attempting to develop standards at the component level.

Digital Document Imaging Systems can generate many storage formats. Although a Digital Document Imaging System is usually coherent and self-integrated, it could be difficult, if not impossible, to transfer a

digital document from that system to a dissimilar system. The main reason why manufacturers are reluctant to embrace common standards is that they have not standardized their own proprietary features, which set their products apart competitively. The computer industry traditionally sells hardware/software interfaces, protocol converters, and so on, to promote connectivity and compatibility rather than embrace new and different standards that might make existing and expensive technology obsolete. These short-term converter/adapter industry solutions frequently provide limited interoperability and integration. Even though an organization can convert all data for exchange if it uses enough computer machine cycles, the costs in time, personnel, and complexity, could be unacceptably high.

Optical scanners make the process of creating digital document images easy, fast, and affordable. The *de facto* standard output format for digital images is uncompressed TIFF—a noncopyrighted format for storing bilevel, grayscale, and color images. TIFF allows variation in its implementation for hardware and software flexibility, with more than 40 variations currently in existence. Optical Character Recognition (OCR)—the process of producing a text file of ASCII codes from images of text—works from scanner output and usually expects TIFF as input. TIFF can be readily compressed, so Digital Document Imaging Systems frequently use TIFF for both exchange and storage. *Be aware that the many variations of TIFFs can hinder, rather that support, data exchange between applications.*

Graphic images and formats are linked to the hardware that displays and outputs them. This fact makes the possibility of low-level (machine-level) standards for images unlikely. Even though no mandated standard for image exchange exists, the Government Open Systems Interconnection Profile (GOSIP), developed by the National Institute of Standards and Technology (NIST), likely will promote the encoding algorithms (data compression) developed by the Consultative Committee on International Telegraphy and Telephony (CCITT)—specifically the Federal Information Processing Standards (FIPS) for Group 3 (FIPS 147) facsimile (fax) transmission, and Group 4 (FIPS 150) high resolution and compression of digital images. Although these two standards are not optimized for a computer's internal use and data representation, they are the formats for exchange of raster-based images.

NIST based GOSIP on Open Systems Interconnection (OSI) standards. Adopted as a Federal standard (FIPS 146) in 1989, GOSIP requires all applicable equipment procured by Federal agencies after August 1990 to comply with the standard.

### 3.3.1.  Document Input

Several standards concern the capture of information and the receipt and subsequent manipulation of this electronically coded data. Cables and interface hardware and software connect image scanners to a Digital Document Imaging System. Manufacturers usually build these scanners to electrical interface standards (RS-232 and IEEE-488) and allow the use of a Small Computer Systems Interface (SCSI) architecture (FIPS 131).

Optical character recognition (OCR) devices can use image scanners to input documents for processing. OCR devices identify images captured by a scanner as letters and numbers by using either a matrix-matching technique or feature-extraction algorithms. OCR devices then convert the input data to a form that word processors, database programs, spreadsheets, or other applications can use. High-performance OCR devices that employ a combination of matrix matching and feature-extraction algorithms are called Intelligent Character Recognition (ICR) devices. Some OCR devices can separate images from text and supply digital images in a TIFF or CCITT Group 4 format. Artificial Intelligence (AI) and neural network technology promise to make OCR devices translate text images more completely and accurately.

FIPS 32-1, 85, 89, 90, and 129 all address the issue of text quality for OCR. FIPS 157 adopts the industry standard—ANSI/AIIM MS44-1988—as a guideline for quality control of image scanners. ANSI/AIIM MS44-1988 includes test targets and procedures for establishing a quality reference and for monitoring scanner quality.

### 3.3.2. Document Storage

Due to the sheer volume of material that organizations need to store, most Digital Document Imaging Systems use optical storage technology. Three types of optical media technologies predominate: Compact Disc–Read-Only Memory (CD-ROM) disks, Write-Once, Ready Many times (WORM) disks, and Magneto-Optical (M-O) rewritable disks. The International Organization for Standardization (ISO) has issued standards for CD-ROM devices (ISO 9960 and 10149). The American National Standards Institute (ANSI) is currently developing standards for WORM and M-O devices (ANSI X3B11).

A fourth technology, Digital Audio Tape (DAT), is expected to be available soon. It is not known, however, if current litigation involving the use of DAT for music recording will have any effect in delaying its availability for use in computer data storage.

Optical disks are currently available in four sizes: 5.25-, 8-, 12-, and 14-inch. The industry itself is in the process of eliminating the 8-inch size. The 5.25-inch size is the industry standard for CD-ROM. WORM has no published standard and is a manufacturer-specific *de facto* standard.

Optical media use one of three current recording formats: Constant Angular Velocity (CAV), Constant Linear Velocity (CLV), or a hybrid CAV/CLV. Due to the different formatting methods used by various manufacturers, optical disks frequently cannot be exchanged from one manufacturer's drive to another—even if the two disks are the same size and media type.

Standards organizations have developed several physical and electrical standards to allow for plug-in connectivity between peripherals (disk drives, printers, scanners, and so on) and a server. Even though the industry is adopting most of these standards, some manufacturers either have not adopted these standards or have adopted them at their own discretion. Two prominent standards are the Electronics Industries Association's (EIA's) physical and electrical standard, RS-232, and American National Standards Institute's (ANSI's) architectural standard, Small Computer System Interface (SCSI).

### 3.3.3. Document Output

Output standards relate to the presentation of output, either printed or displayed online, to the end user. A standardized end-user interface can help enhance productivity, alleviate constant retraining, and avoid processing isolation. No other aspect of standardization affects the user as directly as output standards. FIPS 158 is the acceptance of the X Window system developed by the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts. The National Institute of Standards and Technology (NIST) Application Portability Profile (APP) includes this standard. The X Window system is a vital link to achieve portability and interoperability. X Window–based software will likely be the future standard, particularly for workstations.

Printer manufacturers developed Page Description Languages (PDLs) to produce printed output in a standardized form. The *de facto* standards for PDLs are Hewlett-Packard's Printer Command Language (PCL), used in the LaserJet Series II and compatibles, and Adobe Systems' Encapsulated PostScript (EPS), used in the Apple LaserWriter II NTX and other PostScript-compatible laser printers. Again,

even though PDLs are standardized methods for producing output on printers, they are inappropriate for data exchangeability via networks or telecommunications because the files they produce can be large.

The Portable Operating System Interface for Computer Environments (POSIX) is a recently published standards for operating systems (FIPS 151; IEEE 1003.1). POSIX, which is based on UNIX (a multiuser, multitasking operating system), can run on a variety of computer platforms. Common platform-specific operating systems are Microsoft's Disk Operating System (MS-DOS) for PCs, Digital Equipment Corporation's (DEC's) Virtual Memory Operating System (VMS) for minicomputers, and IBM's Multiple Virtual Systems (MVS) for mainframes.

### 3.3.4.   Document Processing

The ability to perform format conversion improves compatibility and enhances interoperability for image formats. Although text format conversion is usually straightforward, image conversion is more complex and time consuming due to different format types and large-size files. In general, it is better for systems to use a common image format than to rely on format conversion. Software is available to convert raster to vector and vice-versa. In addition, other software, such as Micrografx Inc.'s XPort, can create an intermediate file from which all bidirectional vector conversion takes place.

FIPS 152 adopts the Standard Generalized Markup Language (SGML) ISO standard (ISO 8879). SGML specifies a structure for document representation that can be used for applications of generic coding and markup. It provides a coherent and unambiguous syntax for describing the construct and context of a document. This standard for generically tagging digital document components allows for the consistent exchange of documents. As of August 1989, all Federal agencies developing or acquiring in-house publishing systems (or other similar document-processing systems) must assure that they comply with the SGML.

FIPS 127 mandates that ANSI SQL 89—Structured Query Language (SQL)—is the standard interface for any database management system (DBMS) purchased by Federal agencies. SQL is a language for manipulating relational database information. FIPS 127.1 adds embedded SQL (SQL made a part of a compiled program's code with an assurance of portability) to the existing SQL requirement as well as a optional requirement for referential integrity (which ensures that each data item in a database belongs to a defined group). Standard SQL could be the only near-term solution to provide distributed database client/server interoperability.

Digital Document Imaging Systems can store images in one of several file formats. A *de facto* format is Aldus Corporation's Tagged Image File Format (TIFF) even though its implementation of the standard varies among vendors. Another common file format is ZSoft's PCX. Most other vendors' offerings use proprietary file formats.

### 3.3.5.   Communications

The large size of digitized images has a significant impact on communications. Even though optical disks in Digital Document Imaging Systems need access to networks, many network operating systems do not fully support this access. Digitized images, which can need a wide bandwidth for high-volume or fast transmission, are often supported through a T-1 circuit. Communications networks also need an appropriate protocol and large packet size to transport efficiently image files. Some Digital Document Imaging Systems are designed to route image traffic over an image network. This image network is designed to be separate from other text networks, particularly those used to support an online transaction processing (OLTP) system.

Compression is a process of applying an encoding algorithm that allows the same amount of information to take up less space (a smaller file) and therefore reduces storage and communications costs. CCITT Group 4 (FIPS 150) is the current encoding standard for two-dimensional raster images. The U.S. Department of Defense (DoD) supports CCITT Group 4 in its Computer-aided Acquisition and Logistics Support (CALS) initiative as part of MIL-R-28002. The purpose of CALS is to provide a consistent method of exchanging text and images among multiple computer hardware and software platforms. Any vendor or contractor wishing to do business with the DoD, must be CALS compliant by 1992.

Digital Document Imaging Systems currently can use one of several encoding algorithms. However, CCITT Group 3 or Group 4 compression/decompression standards are by far the most common. Unfortunately, vendors do not offer uniform implementations of either Group 3 or Group 4. A raster image compressed by one vendor's Group 4 implementation does not necessarily decompress correctly when using another vendor's Group 4 implementation. The method used to perform compression/decompression can be either hardware- or software-based.

Facsimile (fax) transmission is a telecommunications method for sending images using modems and telephone switching systems. A fax machine processes a fax transmission in real-time, with required document queuing taking place at the sending location. The availability of low-priced hand-held scanners has increased the popularity of PC-based fax boards. A fax board is a logical method to transmit electronically an occasional page image from a Digital Document Imaging System.

Fax transmission is similar to electronic mail (E-mail)—a store-and-forward, send-only message service. Fax devices are graphics-oriented and usually do not send ASCII-coded documents. Fax machines complying with the CCITT Group 3 standard (FIPS 147) do not have sufficient resolution for OCR devices to convert the received printed image pages to an ASCII-coded document. Fax machines that support CCITT Group 4 (FIPS 150) makes this conversion feasible, but these machines require expensive communication bandwidths. Future implementations of Integrated Services Digital Network (ISDN) could help make Group 4 fax more cost effective.

Standards organizations and industry have established several standards in communications. The network architecture standard currently pursued most is the International Organization for Standardization (ISO) Open Systems Interconnection (OSI) reference model. Implementations of this model in Local Area Networks (LANs) include Xerox's Ethernet and IBM's Token-Ring topologies; both are ISO standards.

### 3.3.6.    Standards in the Government Open Systems Interconnection Profile (GOSIP)

Practical considerations regarding document management requires using image technology before some applicable industry or Government standards are firmly established. After determining the need and believing that imaging technology can help fill that need, the next concern is deciding which imaging technologies to use and how to integrate them into the workplace. To maintain a degree of consistency and operational compatibility, a thorough understanding is necessary of the standards that should be part of the specifications for a Digital Document Imaging System.

When considering such a system, bring telecommunications experts into the planning stages so the system can meet the special needs of image processing and yet be able to conform to long-term objectives of data exchangeability and system interoperability. An important set of standards to consider regarding the interoperability of systems is the Government Open Systems Interconnection Profile (GOSIP).

GOSIP is a Federal standard (FIPS 146), issued August 24, 1988, that defines a common set of data communications protocols which enable systems developed by different vendors to interoperate and enable the users of different applications on the systems to exchange information. GOSIP has been established as a single networking architecture for all Federal agencies. GOSIP offers a partial solution to image processing interoperability. The Federal Government has mandated that, when acquiring network products, services, and communication systems after August 1, 1990, Federal agencies must ensure their equipment procurements meet initial GOSIP specifications. The GOSIP standards that apply to Digital Document Imaging Systems apply equally to data and text systems.

As GOSIP standards continue to emerge, the National Institute of Standards and Technology (NIST) gives each revision a new version number. Thus, the first release of GOSIP is version 1; subsequent releases will follow numerically. NIST indicates it will issue new versions at no less than one-year intervals.

Subsequent GOSIP releases will include the following specifications:

- Office Document Architecture/Office Document Interchange Format (ODA/ODIF)
- Integrated Services Digital Network (ISDN)
- File Transfer, Access, and Management (FTAM)
- Fiber-optic Distributed Data Interface (FDDI)
- Message Handling System (MHS) extensions

## *Office Document Architecture/Office Document Interchange Format (ODA/ODIF)*

Implementation of the open systems described by GOSIP and OSI will enable the movement of a document to a new environment where it can be understood and processed as if it were in its original system environment. The ODA/ODIF standard (ISO 8613) allows for the exchange of compound documents (documents containing text, facsimile, and graphics) generated by diverse office products. The exchange can be made by communications or exchange of storage media. Images exchanged as documents in this manner would fall under this standard applicability.

## *Integrated Services Digital Network (ISDN)*

Designed to be an all-digital network, ISDN accommodates simultaneous transmission of voice and data over existing telephone circuits. Yesterday's circuit-switched analog circuits are rapidly disappearing in favor of today's packet-switched digital circuits. Digital circuits provide greater capacity than analog circuits due to increased bandwidths (frequency range), faster transmission rates, and decreased distortion (signal-to-noise ratio). ISDN supports CCITT X.25—the interface standard for connection of Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE). ISDN holds promise for telephone (desktop) conferencing, where participants can use communications lines to exchange data and electronic documents while simultaneously having a voice discussion. In summary, ISDN is expected to offer high-speed, dial-up access to many computer networks and will allow rapid exchange of documents, data, and, possibly, digital images.

## *File Transfer, Access, and Management (FTAM)*

Designed to be a set of control and file-transfer services between networks, FTAM is a standard (ISO 8571) that includes support for remote record access, including addition, modification, and deletion of data and records. FTAM supports digital (binary) and text file formats and does not require the user on one network to know the detailed file characteristics of the manipulated file on the second network.

*Fiber-optic Distributed Data Interface (FDDI)*

FDDI is a standard (ANSI X3.139) that provides specifications for 100 megabits per second (Mbps) data transfers for Local Area Networks (LANs) using fiber-optic cables.

*Message Handling System (MHS)*

A component of the CCITT X.400 recommendations, MHS concerns electronic messaging, or electronic mail (E-mail). The CCITT X.400 recommendations support messaging functions to include exchange of digital (binary) and text files, electronic carbon copies, directed replies, and prioritized messages. The X.400 recommendations establish the interface between proprietary electronic message networks, such as MCI Mail. The CCITT X.500 recommendations define the directory services that X.400-compliant systems must use to determine the proper destination address of a message's intended recipient.

As of August 1990, Federal agencies that acquire communications network products, services, or systems (including systems connected to networks) must specify compliance with GOSIP (FIPS 146) standards. Pay special attention to the expanding requirements and standards specifications issued by the National Institute of Standards and Technology (NIST).

The major expected benefit of standardization is the ability to move with the future and not be left behind in a vendor's obscure file format, processing method, or system. Even if a proprietary system offers processing speed and special features, it could produce data that is not convertible. The time or money saved by choosing a proprietary system now can be of little significance later when exchangeability, interconnectivity, and interoperability become essential to maintain the operation.

# 4. TECHNICAL SUMMARY AND GENERAL GUIDELINES

In preceding sections, discussions focused on the following: digital images as essentially a new data type that can be effectively applied to meet selected functional needs; long-term objectives regarding data exchangeability and system interoperability; and digital imaging equipment, formats, and standards. This section contains a brief summary of digital imaging technology, general guidelines for effective planning and implementation of a Digital Document Imaging System, and a brief list of tips to follow and traps to avoid when considering such a system.

## 4.1.   SUMMARY OF DIGITAL IMAGING TECHNOLOGY

The digital imaging marketplace is growing and changing rapidly. As with other maturing technologies, prices are beginning to drop and, in all likelihood, will continue to fall. However, proprietary interests continue to cloud the technology, making assessments difficult and possibly incomplete, and making standards in segments of the market elusive. Although few standards exist today, managers can apply sensible strategies to minimize the effects of this condition. In addition, a largely nontechnical sales force services this market, which can make it difficult to assess all capabilities and shortcomings of individual products. The predominance of vertical market products, such as standard desktop-publishing applications, requires system integration for atypical applications.

Digital imaging can be expensive and labor intensive if it is not applied carefully and surrounded with sufficient processing capability, management software, and communications support. Although digital imaging is an important new technology, it is still complex and has a small, but growing, experience base upon which to draw for assistance. To use digital imaging technology wisely could require redesigning the workplace and its infrastructure.

The Federal Government now requires systems to have adequate communications, security, data integrity, system reliability, backup, and disaster recovery capabilities. In response to needs, these capabilities can be augmented with the capability to capture, store, manage, retrieve, and display digital document images.

## 4.2.   GENERAL GUIDELINES

The following guidelines are offered as a basic approach to take when considering a Digital Document Imaging System. Because each situation is different, these guidelines can suggest only general directions.

- Consider Open-Architecture Systems over Proprietary Systems
- Analyze and Plan Carefully
- Rely on a Systems Approach to Planning and Implementation
- Specify Workflow Requirements in Detail
- Examine Document Characteristics Closely
- Approach Retrospective Conversion Cautiously
- Carefully Assess Communications Requirements
- Reduce Uncertainties through Risk Assessments and Prototyping
- Manage User and Management Expectations

*Consider Open-Architecture Systems over Proprietary Systems*

To achieve the objectives of data exchangeability and system interoperability, an open-architecture design is preferred. Open systems can be used to take advantage of future technological advances. They also allow component and software exchange and can provide a possible path to integrated information systems of the future.

Wherever possible, choose imaging system components that meet industry, Government, and NASA standards. Study the standards and understand the extent of compliance with individual standards offered by different vendors. Consider proprietary products only when no applicable standard is available or when the standard does not provide the functionality required.

In today's environment, Digital Document Imaging Systems based on an open architecture usually have the following characteristics:

- Client/server architecture
- Sophisticated applications development environment, including fourth-generation programming tools, function libraries, and C-based application programming interfaces
- Off-the-shelf computers and Local Area Networks (LANs)
- Fully distributed relational database management systems (DBMSs) with an ANSI-standard Structured Query Language (SQL) interface
- Standard communications protocols and document formats

As an example, a Digital Document Imaging System that uses the following component specifications will provide support for an open architecture:

- Integrated IBM 3270 windowing capability
- C programming language
- Microsoft Windows with Dynamic Data Exchange (DDE)
- CCITT Group 3 and Group 4 image compression and transmission
- Tagged Image File Format (TIFF)

*Analyze and Plan Carefully*

The practical functionality, costs, and promised benefits of digital image management require careful analysis, and systems implementation requires careful planning. This analysis must begin with a clear assessment of user needs and be driven throughout by user needs. Avoid becoming captivated by the mystique of digital imaging—it is merely a tool for meeting users needs. Success means planning and selecting the system that best meets those needs. To participate in any future information system that NASA might implement, develop strategies for achieving data exchangeability and interoperability before selecting a system.

*Rely on a Systems Approach to Planning and Implementation*

Apply a structured, top-down approach to systems planning and implementation. As with all successful integrations of new technologies, this approach should begin with an examination of users needs and provide for the continual assessment of technical and functional approaches against those defined and changing needs.

Following an assessment of user needs, examine alternative methods for meeting these needs. Evaluate the marketplace for both text and image management systems. Assess digital image management technology and enabling technologies. Review the results of these efforts against the defined user needs and formulate both short- and long-term functional and technical strategies. Next, assess specific system requirements, such as capacity, access, retrieval, and communications. In addition, consider various system architectures. Once the system requirements and an architecture are determined, alternative acquisitions strategies can be evaluated. The system or subsystem then can be designed in detail, developed, and implemented. It can be helpful to follow a life-cycle methodology throughout this process. A good example is the methodology detailed in the *AIM Program Technical Manager's (PTM) Guidebook* issued by the NASA Office of Information Resources Management (IRM), Automated Information Management (AIM) Program Office.

### *Specify Workflow Requirements in Detail*

Workflow represents the stages required for a particular task to be completed in an organization. It includes the flow of information and the impact that information has on individuals and processes involved. In the workflow analysis, include not only document-handling and records management tasks, but also accounting and other functions, such as customer services.

Workflow analysis can reveal how an organization's current infrastructure uses information, in all of its forms. This understanding is necessary to determine which technologies can provide the most benefit when applied to specific needs. Digital Document Imaging Systems can aid in automating information management. When applying digital imaging, however, simply automating an existing process does not necessarily produce cost-effective results.

Workflow analysis will help to reveal system requirements, such as the frequency and volume of document images that must be transferred and processed at each stage. After determining the required imaging system workflow, investigate the level of integration that the system will require. Implementing imaging systems can produce measurable gains in productivity, but only if the level of integration significantly improves the way an organization operates. With the workflow quantified and the level of integration characterized, the type and amount of effort required to implement the technology can be specified.

A properly integrated, efficient imaging system can offer improved workflow, better document control, faster retrieval, and overall enhanced service. Under some circumstances, savings are possible through a smaller work force, reduced use of paper or microfilm, and/or lower storage costs. However, digital imaging technology cannot be justified simply as a less costly way of maintaining current operations. When used on a large-scale, the technology is expensive and could require a redesign of the workplace and its infrastructure. When applied to the most appropriate workflow, however, digital imaging technology can transform how information is stored, retrieved, and distributed, while providing both tangible and intangible benefits.

### *Examine Document Characteristics Closely*

For the successful application of digital imaging technology to meet document-handling needs, the characteristics of the documents that require automated management should be examined. Many features contained on a page in a document can affect the quality of a digitally scanned image as well as affect the success of OCR devices in accurately identifying text.

Although the technologies currently involved in digital document imaging continue to improve, they are sensitive to the following characteristics: size, finish, weight, and thickness of paper; page composition and layout; and the use of color, line art, halftones, or photographs. If OCR capability is being considered as part of a document-management solution, be aware of the variety of type size and fonts contained in the document pages, complexities of page composition, and the use of symbolic text, special characters, and scientific/mathematical notation. After determining the frequency with which these features occur in the documents planned for digital imaging, strategies and specific plans can be developed for the most effective use of specific components of digital imaging technology. For example, if a document contains many pages of mathematical equations and symbols, those pages might best be stored as a digital image only without attempting to run them through an OCR device.

## Approach Retrospective Conversion Cautiously

Retrospective conversion involves converting existing documents into digital images. Documents created before the implementation of a Digital Document Imaging System can be in several forms (film, magnetic, paper, optical, and so on) and could require separate, unique conversion processes. Methods used to convert different types or parts of former documents also can vary. Successful conversion takes extensive planning and a comprehensive understanding of both the old and new systems. Be aware that conversion can be expensive, labor intensive, and sometimes produce less than optimal results.

Alternatives to full retrospective conversion include the following:

- limited conversion
- maintaining older systems and inputting subsequent documents on the newer system
- development of hybrid Digital Document Imaging Systems

## Carefully Assess Communications Requirements

A typical file size for an uncompressed digital image of a single 8.5-by-11-inch text page is about 1MB. To avoid unacceptable speed and performance levels, this bit-mapped image file can be compressed to as little as 50KB. Even when compressed, this file would still be much larger than the 2KB file size for the conventionally coded ASCII representation of the same text page. Both communications equipment and available technology require special attention because digital images have such large file sizes. High-volume image file transfers could require careful use of queuing models and simulations. Local- and long-distance communications requirements should be thoroughly analyzed and specified. In addition, the availability and cost of the telecommunications services required to meet specified requirements. Because telecommunications is a rapidly changing field, telecommunications experts should be consulted to help anticipate and evaluate new telecommunications services.

## Reduce Uncertainties through Risk Assessments and Prototyping

Reducing uncertainty is an important process for the successful implementation of a DDIS project. Identify elements that have high risk factors. Appropriate steps can be to reduce risk by obtaining and assessing additional knowledge or identifying alternative lower-risk solutions. Performing additional system analysis, or designing and testing a prototype system, can also lower risk.

A prototype is a realistic model used to test the basic design and operational features of a proposed system before developing a production-level system. Prototyping can be a cost-effective, low-risk method for evaluating a system's functionality and practicality. Efficient tools are essential in prototyping and contrib-

ute to the final design. Prototypes can aid in development of production-level systems through refining requirement definitions, clarifying user interfaces, and uncovering potential problems.

With the possible exception of a small stand-alone system, forming a project team is beneficial in determining actual needs and developing appropriate solutions. This team should be comprised of end users, document-management experts, MIS/ADP personnel, and telecommunications experts. End users should be active in most phases of the development effort.

### *Manage User and Management Expectations*

Expectations of digital imaging technology and its appropriate application by both end users and managers can be unrealistically high. Too often, these individuals view digital imaging technology as the ultimate cure for all problems—it is not. Only if the application of this technology is carefully planned and implemented can the resulting system provide the anticipated results.

Additionally, expectations can also be negative. It is possible to phase-in this new technology gradually so as to cause a natural evolution, rather than a sudden revolution, to take place within an organization. Potential problems can arise when users are faced with replacing traditional paper documents with electronically displayed information. It is important to give users an opportunity to provide design and implementation feedback. By doing so, users will feel they are actively participating in determining major changes to their work environment.

## 4.3. PLANNING/MANAGING DIGITAL DOCUMENT IMAGING SYSTEMS

Below are some suggestions on approaches to take and to avoid when considering a Digital Document Imaging System. Remember that each situation must be considered individually.

### 4.3.1. Helpful Tips

- Confirm the validity of sales representatives' claims regarding the technical capabilities and functionalities of equipment
- Conduct a thorough systems analysis
- Know the needs and problems to be addressed by Digital Document Imaging Systems
- Design the system to fit user needs
- Visit and analyze similar installations of Digital Document Imaging Systems
- Set realistic, conservative goals and inform both management and users of those goals
- Calculate the time, labor, and cost involved with retrospective document conversion
- Involve management since Digital Document Imaging Systems can affect workflow
- *Consider a prototype system to test working assumptions*
- Implement Digital Document Imaging Systems in manageable phases or stages
- Plan for cabling and installation costs
- Adequately size the system to allow for system growth
- Insist on an intuitive interface for increased productivity and ease of use
- To ensure compatibility, insist on hardware and software that comply with FIPS regulations
- Allow for unintended consequences—they happen

### 4.3.2. Potential Traps

- Forgetting the needs of major constituents who are not users
- Underestimating the cultural changes that new systems can have on users
- Assuming Digital Document Imaging Systems are off-the-shelf, do-it-alone projects
- Believing Digital Document Imaging Systems are cure-alls for every information system problem
- Overlooking the requirements for training that users will need to be able to use a new system
- Assuming that even stand-alone Digital Document Imaging Systems manage themselves
- Undervaluing the time and cost of integrating existing equipment into Digital Document Imaging Systems
- Implementing technology for technology's sake instead of addressing application needs

# 5. APPENDICES

The five appendices that follow provide material to supplement the body of this guide.

*Appendix A*, on page 33, lists Digital Document Imaging System component specifications and, where relevant, details the applicable standards, options, known problems, recommendations, and comments.

*Appendix B*, on page 43, contains detailed descriptions of organizations that issue applicable standards as well as a list that compiles these standards by organization.

*Appendix C*, on page 57, is a cross-reference to two groups of relevant standards: Federal Information Processing Standards (FIPS) and Government Open Systems Interconnection Profile (GOSIP).

*Appendix D*, on page 65, lists additional resources to consult for further information, including printed material, conferences/training seminars, and consultants/research groups.

*Appendix E*, on page 69, contains a glossary of terms used in digital imaging technology and the references used to compile this glossary.

# A. COMPONENT SPECIFICATIONS

This appendix contains specifications of components that compose a Digital Document Imaging System. Principal features of each component are described along with applicable standards. Advice is offered in the form of comments and recommendations as well as listing the known problems or pitfalls of each component. Be aware that the information regarding these components is of a time-sensitive nature. Vendors are constantly incorporating new technology into existing products or are offering entirely new products. Figure 3 below shows the different components of a Digital Document Imaging System and their relationships to each other.

**Figure 3.** *Relationships of Digital Document Imaging System Components*

PRECEDING PAGE BLANK NOT FILMED
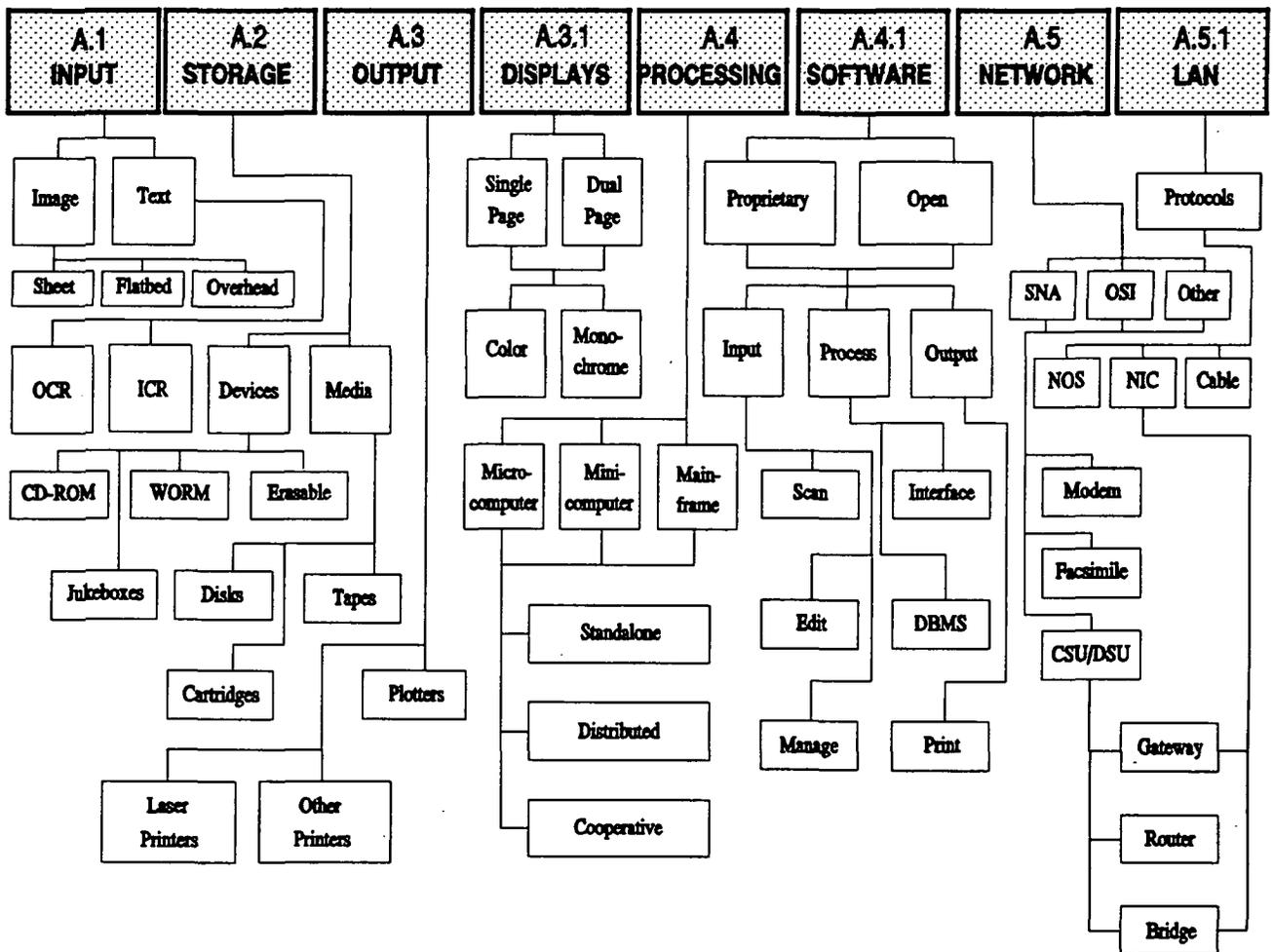
**SUBSYSTEM:  A.1 INPUT**

**COMPONENT:** Graphic Scanners

**FUNCTION:** Convert hard copy image to digital image

**PRINCIPAL TYPES:**
- Line art (bilevel)
- Continuous tone (grayscale)
- Hand-held
- Sheetfed
- Flatbed

**APPLICABLE STANDARDS:**
- ANSI/AIIM MS44 1988, FIPS 157 for recommended quality-control practices
- IEEE RS-232C, PC-bus, parallel, and SCSI architecture for interface

**TYPICAL ISSUES:**
- Resolution range
- Image quality
- Dithering, grayscale levels
- Scan speed
- Dynamic thresholding
- Host computer compatibility
- Brightness, contrast, scaling, and image area
- Optical Character Recognition
- Template recognition
- Feature recognition
- Color support
- Common graphics formats supported (TIFF, PCX, EPS, PIC, IMG, and so on)
- Media handling capability
- Scanner software integration
- Price, vendor's record, and support

**OPTIONS:**
- Automatic sheet feeders
- Specialized software

**PITFALLS/KNOWN PROBLEMS:**
- Make sure scanner outputs compatible format
- Moiré patterns caused by dithered halftones
- Image memory requirements are large
- Documentation can be insufficient
- Match interface scanner to host processor

**RECOMMENDATIONS:**
- Match device to needs
- Integrate scanner to processor
- Check not only speed of scan but scanner's duty cycle
- If a grayscale scanner is needed, get at least a 64-shade capability
- Get at least 12 halftone patterns, brightness, contrast, and scaling settings
- The interface of the scanner needs to be matched to that of the host interface
- The data format, usually TIFF or raw bits, needs to be matched to the requirements of the image-processing software

**COMMENTS:**
- Consider out-sourcing large backfile conversions to service bureaus
- Service and on-site support are key elements in making a choice of scanners
- Generally, a screen resolution of 75 dpi is sufficient for viewing an image of printed text
- Generally, a resolution of 300 dpi is recommended if OCR is planned
- Consider barcoding for document control when hardcopy originals must be kept
- Systems that allow OCR on previously scanned images, provide greater workflow flexibility
- Digital scanning produces large files; an ASCII representation of a page of text usually requires 2KB of storage where a digital image file of that same page usually requires 1MB of storage before compression

**SUBSYSTEM:  A.2 STORAGE**

**COMPONENT:** Storage Devices

**FUNCTION:**  Provide nonvolatile electronic storage of digital images

**PRINCIPAL TYPES:**
- Magnetic disks
- WORM disks
- CD-ROM
- Rewritable disks
- Magnetic tape

**APPLICABLE STANDARDS:**
- Winchester       – Magnetic
- ESDI             – Magnetic
- ANSI X3B11       – WORM and Rewritable
- ISO 9660         – CD-ROM
- ISO CCS format   – Rewritable
- SCSI-1, SCSI-2   – All
- DDS, Data/DAT    – DAT

**TYPICAL ISSUES:**
- Storage cost per megabyte
- Formatted capacity
- Transfer rate of access device
- Initial and average seek times
- Physical storage requirements of device
- Interfaces and computers supported
- Network operating systems supported
- Life and cost of media

**OPTIONS:**
- Jukeboxes
- Media with higher capacity
- Special interfaces

**PITFALLS/KNOWN PROBLEMS:**
- Different media have different environmental problems
- Compatibility problems
- Useful capacity is sometimes much less than theoretical capacity due to software and/or hardware limitations
- Provide for backups and audit trail if necessary
- Account for full life-cycle storage requirements of the system

**RECOMMENDATIONS:**
- Buy hardware and software based on proven record
- Storage devices should be an open-systems component
- Try to avoid a sole source for media needs
- Write specifications for space and performance requirements
- Look for strong warranty for protection

**COMMENTS:**
- Document storage is one component that requires precise integration to the processor(s) for an efficient system
- If future needs include a jukebox, keep this requirement in mind when selecting optical drives
- Optical drives require a controller which is usually supplied, an interface such as SCSI and software so that it is known to the operating system and the image applications
- An open-systems optical device allows for retention of data when the host processor is replaced

**SUBSYSTEM:   A.3 OUTPUT**

**COMPONENT:** Printers and Output Devices

**FUNCTION:**     Produces hard copy image to a digital image

**PRINCIPAL TYPES:**
- Laser
- Ink-jet
- Color
- Imagesetters
- COM
- Facsimile

**APPLICABLE STANDARDS:**
- PCL
- EPS
- IPDS
- DDL

**APPLICABLE INTERFACES:**
- Direct video interface
- RS-232C or RS-422 serial
- SCSI-1 or SCSI-2
- Centronics or IBM parallel
- Other manufacturers' specific interfaces

**TYPICAL ISSUES:**
- Resolution
- Memory
- Page print mechanism (engine)
- Print speed in graphic modes
- Paper handling
- Automatic duplexing
- Paper sizes
- Emulations supported
- Font mechanism
- Duty cycle

**OPTIONS:**
- Paper trays
- Additional memory
- Font cartridges or font software

**PITFALLS/KNOWN PROBLEMS:**
- Printer either over- or under-engineered for the task
- Inadequate preparation given to the location of the printer
- Manufacturers' quoted text speeds are not always reliable

**RECOMMENDATIONS:**
- 300-dpi laser printers are a common choice for Digital Document Imaging Systems
- Laser printers with direct video interface (Ricoh or Canon engine) are able to print graphic images at near text speeds with the proper host imaging board and interface

**COMMENTS:**
- High-resolution color printers, although much slower than their single-color counterparts, are now available for under $10,000
- Image software and the printer chosen must be compatible to have an operable system
- Proper configuration can enhance a printer's versatility and improve productivity
- Consider multitasking several low-cost laser printers instead of using one high-speed unit, which can be a single source of failure

SUBSYSTEM:   A.3.1   Displays

COMPONENT: Monitors

FUNCTION:   Provide visual images, usually on monitors, of digital images

PRINCIPAL TYPES:
- Single- or dual-page
- Analog, TTL, or composite

APPLICABLE STANDARDS:
- Various industry standards

TYPICAL ISSUES:
- Screen size, display area, case size, footprint
- Resolution
- Pixel size, dpi, video bandwidth
- Number of grayscale shades supported
- Compatibility modes
- Required and supported display adapter boards
- Software drivers supporting common applications at higher resolutions

OPTIONS:
- Swivel-and-tilt stand
- Antireflective coating or screen
- Color
- Monochrome

PITFALLS/KNOWN PROBLEMS:
- VDT flicker can cause eyestrain
- Monitors require good airflow to dissipate heat properly

RECOMMENDATIONS:
- Most applications use monochrome monitors for viewing images, but color is helpful for process quality control
- Image capture and quality control functions can be helped with a dual-page monitor if the application can take advantage of this feature
- Single-page monitors are adequate for display when there is no need for Dynamic Data Exchange (DDE) of documents and system monitoring

COMMENTS:
- Basically, what you see is what you get (WYSIWYG)
- Do not undervalue this important human interface with the computer
- High-scan frequencies reduce flicker and lessen user fatigue
- A small dot-pitch ($\leq.31$mm) is essential for sharp images

**SUBSYSTEM:  A.4 PROCESSING**

**COMPONENT:** Computer Base Units and System Architectures

**FUNCTION:**  Provides digital processing power to service digital images with image access, compression/decompression and manipulation

**PRINCIPAL TYPES:**
- Microsoft    – MS-DOS
- IBM          – OS/2
- AT&T         – UNIX System V/AOE
- Apple        – Macintosh System
- UNIX         – Open Systems/OSF1
- DEC          – VAX/NAS
- HP           – NewWave
- IBM          – 370/MVS/SAA
- Wang         – OPEN/Architecture
- Others, including Bull/DSA, CDC/NOS, Data General/DAA, and NCR/OCCA

**APPLICABLE STANDARDS:**
- Still ruled by the marketplace
- POSIX software standards for the operating system

**TYPICAL ISSUES:**
- Processor's speed (in MIPS)
- Bus transfer rates
- Amount, type, and speed of memory
- Open-systems architecture
- Availability of same architecture across different platform sizes
- Expandability
- Upgradeability
- Throughput

**OPTIONS:**
- Usually limited only by budget constraints
- Operating systems and software supported
- Image processing boards
- Second party availability of basic components and ancillary devices

**PITFALLS/KNOWN PROBLEMS:**
- Performance of processor alone is not a valid indicator of system quality or reliability
- Requires analysis to determine memory space requirements

**RECOMMENDATIONS:**
- Try to avoid sole-source equipment or vendor
- Second-party availability of basic components and ancillary devices will increase availability and usually reduce cost
- POSIX compliance (FIPS, NIST, and GOSIP) makes it difficult for the operating system not to be a UNIX/POSIX version
- The cost-effective architecture for UNIX is Reduced Instruction Set Computer (RISC)

**COMMENTS:**
- Manufacturers' proprietary schemes often protect their product base for competitive advantage to the detriment of their users
- While progress has been made, interoperability and connectivity are largely realized with third-party bridges that require additional hardware and/or software
- If all other factors are equal, bridges and other fixes imply performance degradation
- Compatibility requires exacting physical, electrical and software identity
- Insist that claims of compatibility are backed up with conformance testing
- Imaging does not require a special-purpose processor, but does require high-end graphics equipment and capability

SUBSYSTEM:  A.4.1    Software

COMPONENT: Image Strategies

FUNCTION:    Provide means to control and manage an image system

PRINCIPAL TYPES:
- Proprietary schemes
- SQL/DBMS integrated
- Hypertext/Hypermedia
- Stand-alone
- Image servers to a mainframe/minicomputer host
- Client/server image systems

APPLICABLE STANDARDS:
- ANSI
- OSF
- CALS
- ITU/CCITT
- IEEE
- NIST/FIPS
- OSI/GOSIP
- Manufacturers' and vendors' standards

TYPICAL ISSUES:
- Multiple methods of retrieval and search
- Interoperability
- Application development
- Audit trail
- Image manipulation
- Data security

OPTIONS:
- GUIs
- Integration with voice and video
- Alternate media storage

PITFALLS/KNOWN PROBLEMS:
- All problems that can go wrong with software
- Beware deals that sound too good to be true... they are
- Watch out for exorbitant LAN version costs
- System quickly outgrown due to inadequate analysis

RECOMMENDATIONS:
- Consider vendors track record, total service offerings, and technical ability
- Buy an open-architecture system unless special needs can be met only by a proprietary system
- Build a system in stages and test in parallel before cutting over
- Write the RFP as a functional specification with requirement for the vendor's solution to meet the following: compliance with standards, modularity in software and hardware, and appropriate connectivity options to other open-architecture systems for the future

COMMENTS:
- Rely on a systems integrator and image analyst to provide solutions for a complete life cycle that is tolerable in this complex technology
- Investments in imaging technology usually provide more in the way of intangible benefits to the users and less in payback to the organization and therefore are not easily cost/benefit justified
- Most image systems are installed to handle a specific problem or overwhelming need that serve as the justification of the purchase instead of a measured cost/benefit decision

SUBSYSTEM:  A.5 NETWORKS

COMPONENT: Communications Networks

FUNCTION:    Provide interconnections for nodes,
             LANs, or systems by communications
             channels, usually with common-carrier
             circuits

PRINCIPAL TYPES:
- TCP/IP
- IBM/SNA
- ISDN
- X.25
- DECnet

APPLICABLE STANDARDS:
- OSI, GOSIP, and FIPS
- AT&T, IBM, and other manufacturers
- FCC
- CCITT
- DDS

TYPICAL ISSUES:
- Traffic capacity
- Network architecture
- Interface standards
- Security
- Bandwidth
- Initial and on-going costs
- Communications software
- Network topology
- Network monitoring and management
- Connectivity and interoperability

OPTIONS:
- Public Data Services (PDS) versus private cus-
  tomization

PITFALLS/KNOWN PROBLEMS:
- Insufficient bandwidth for images causing very
  slow transmissions
- Most currently available networks are unsuit-
  able for transporting images

RECOMMENDATIONS:
- Carefully evaluate non-network communica-
  tions alternatives for other methods of image
  distribution

COMMENTS:
- GOSIP gives Federal users little choice but to
  be OSI-compliant
- Facsimile offers a compromise solution for
  some image transportation needs
- ISDN potentially offers the needed bandwidth
  and innovative technology required to help
  resolve the communication problems of images
  as a circuit-switched solution
- A packet-switched network solution for a LAN
  of images will require the bandwidth of T-1
  lines or greater
- The increase in popularity of video-conferenc-
  ing could help provide the affordable, real-time
  communications service that images need for
  long-distance travel

**SUBSYSTEM:  A.5.1    Local Area Networks**

**COMPONENT:** Network Topology, Cabling, Adapter
Cards, and Software

**FUNCTION:**    Provides data communications for con-
nected workstations within a small
area at medium to high data rates and
does not use common carrier circuits

**PRINCIPAL TYPES:**
- Novell NetWare
- Microsoft OS/2 LAN Manager
- Banyan VINES
- Sun NFS
- Apple Appletalk
- IBM PC LAN
- AT&T StarLAN

**APPLICABLE STANDARDS:**
- OSI, GOSIP, and FIPS
- Various manufacturers' standards
- IEEE 802-series standards

**TYPICAL ISSUES:**
- Protocols
- Topology
- Security
- Transfer rate and throughput
- Connectivity
- Wiring types and costs
- Network interface cards
- LAN monitoring and management

**OPTIONS:**
- Sharing of devices among multiple users
- Gateways and bridges are available to many
computer systems and services

**PITFALLS/KNOWN PROBLEMS:**
- Bandwidth of LAN is too narrow for on-
demand transport of images, can cause fre-
quent system crashes
- File server of LAN operating system does not
recognize or interface with optical storage
devices

**RECOMMENDATIONS:**
- Use the experience of qualified experts
- Purchase for performance with adequate re-
serve capacity and not technology hyperbole
- A LAN should be part of an overall workflow
solution
- Plan for technical and administrative manage-
ment of nontrivial LANs
- A current solution to the demands of an image
LAN is the deployment of dual LANs—that is,
an image LAN separate from the text LAN

**COMMENTS:**
- In contrast to text applications, imaging puts
heavy demands on a LAN
- Imaging requires that a LAN for transporting
images be engineered specifically for that task,
especially if fast response time is needed
- Look for the high-end network (LAN) operat-
ing systems to provide compliance with GOSIP
- Expect many systems to run concurrent proto-
cols to handle the diverse requirements of
interoperability until more systems become
OSI-compliant

# B. STANDARDS-ISSUING ORGANIZATIONS

By its very nature, the information contained in this section is dynamic. Not only are new organizations being created, but also existing ones are changing or revising their standards. Every effort has been made to see that this information current and correct.

## B.1. ORGANIZATIONS—DETAILED DESCRIPTIONS

Parts of the information about the organizations in sections B.1.1 through B.1.7 are taken verbatim from either the *Encyclopedia of Associations* (1990) or the *Encyclopedia of Associations: International Organizations* (1990).

### B.1.1. Association for Information and Image Management (AIIM)

AIIM was founded in 1943. AIIM operates with a staff of 25 and a budget of $4,000,000 with a membership of 6,500. AIIM members are manufacturers, vendors, and professional users of information and image management equipment, products, and services. AIIM holds meetings for trade members and companies, presents awards, medals, fellowships, and scholarships for distinguished service to industry. AIIM was formerly the National Microfilm Association; (1983) National Micrographics Association. AIIM publishes the following materials:

1. *Buying Guide* (annually)
2. *FYI/IM* newsletter (10 times a year)
3. *INFORM* magazine (10 times a year)
4. *Resource Center Index* semiannually)
5. *Resource Reports* (periodically)

Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
301/587-8202
Sue Wolk, Executive Director
Marilyn Courtot, Director, Standards and Technology

### B.1.2. American National Standards Institute (ANSI)

ANSI was founded in 1918. ANSI operates with a staff of 107, and a budget of $8,800,000 and has 1,250 members. ANSI members are industrial firms, trade associations, technical societies, labor organizations, consumer organizations, and Federal agencies. ANSI serves as a clearinghouse for nationally coordinated voluntary standards for fields ranging from information technology to building construction. ANSI gives status as American National Standards to standards developed by agreement from all groups concerned, in such areas as definitions, terminology, symbols, and abbreviations; materials, performance characteristics, procedure, and methods of rating; methods of testing and analysis; size, weight, volume, and rating; practice, safety, health, and building construction. ANSI also provides information on foreign standards and represents U.S. interests in international standardization work. ANSI was formerly (1918) American Engineering Standards Committee; (1928) American Standards Association; (1966) United States of America Standards Institute. ANSI publishes the following materials:

1. *ANSI Reporter*, a biweekly newsletter covering issues that affect the voluntary standards system (ISSN: 0038-9676)
2. *Catalog of American National Standards*, an annual list of 8,000 current ANSI approved standards ($20 for nonmembers)

3.  *Progress Report*, a periodic review of association activities and accomplishments

4.  *Standards Action*, a biweekly newsletter listing newly published standards by ANSI and other organizations, including foreign (ISSN: 0038-9633)

American National Standards Institute
1430 Broadway
New York, NY 10018
212/354-3300; 212/302-1286 (fax)
Telex: 42 42 96 ANSI UI
Donald L. Peyton, President

**B.1.3.    International Telecommunications Union (ITU)/Consultative Committee on International Telegraphy and Telephony (CCITT)**

*Union Internationale des Telecommunications* (International Telecommunications Union) is not a voluntary standards organization but a 160 member formal treaty organization administered by the United Nations. The ITU is chartered to assist nations in coordinating and developing efficient, compatible telecommunications systems.

The ITU has four permanent committees; the Consultative Committee on International Telegraphy and Telephony (CCITT) is of significance to the information processing community, focuses on setting telecommunication standards and maintains a library of more than 20,000 volumes on communications and standards. CCITT was founded in 1956, is headquartered in Switzerland and offers memberships to the following groups (current number of members are shown in parentheses): United Nation members (160), recognized private operating agencies (58 telecommunication service providers), scientific or industrial organizations (146) and major international standards organizations (36).

*To order CCITT documents in the U.S., contact:*

Union Internationale des Telecommunications (ITU)
Place des Nations
CH-1211 Geneva 20
Switzerland
41 22 995111 (phone)

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
703/487-4650

**B.1.4.    Electronic Industries Association (EIA)**

EIA was founded in 1924. EIA has over 300 members and operates with a staff of over 100. EIA members are manufacturers of radio, television, photograph, computer, and industrial electronic products and components. EIA sponsors the annual Consumers Electronics Show. EIA supports a number of committees and is actively involved in the development of standards for electrical and electronic components. EIA publishes the following materials:

1.  *Weekly Report*

2.  *International News* (biweekly)

3.  *Industrial Relations Digest* (monthly)

4.  *Industry Yearbook* (yearly)

5.  EIA standards (periodically)

EIA was formerly the Radio Manufacturers Association (1950). In 1965, EIA absorbed the Magnetic Recording Industry Association. EIA holds several conferences annually.

Electronic Industries Association
2001 Eye Street, N.W.
Washington, DC 20006
202/457-4966

### B.1.5.   Institute of Electrical and Electronics Engineers (IEEE)

IEEE was founded in 1963 and operates with a staff of 500 using 10 regional groups and 170 local groups. IEEE has a membership of 274,000. IEEE members are engineers and scientists involved in electrical engineering, electronics, and allied fields. IEEE conducts lecture courses at the local level on topics of current engineering and scientific interest. IEEE assists student groups, awards medals, prizes and scholarships for outstanding technical achievement. IEEE was formed by the merger of American Institute of Electrical Engineers (founded 1884) and the Institute of Radio Engineers (founded 1912). IEEE publishes the following materials:

1.   *Directory* (annually)

2.   *Proceedings* (monthly)

3.   *Spectrum* (monthly)

4.   *Standards* (periodically)

Institute of Electrical and Electronics Engineers Headquarters     IEEE Standards Organization
345 East 47th Street                                               445 Hoes Lane
New York, NY 10017                                                 P.O. Box 1331
212/705-7900                                                       Piscataway, NJ 08855-1331
Eric Herz, General Manager                                        212/705-7960

### B.1.6.   International Organization for Standardization (ISO)

*Organisation Internationale de Normalisation* (International Organization for Standardization) was founded in 1947. Located in Switzerland, ISO operates with a staff of 130 and supports publications in the languages of English, French, and Russian. ISO develops and publishes international standards to facilitate the exchange of goods and services and to foster mutual cooperation in intellectual, scientific, technological, and economic spheres of endeavor. ISO maintains a 6,000 volume reference library of national standards of member bodies. ISO maintains 165 technical committees and 2,400 subcommittees and working groups, as well as, sponsoring advisory committees relating to standardization in connection with consumer interests and enveloping countries. ISO publishes the following materials:

1.   *Press Service* (weekly)

2.   *Bulletin* (monthly)

3.   *Activities Report* (annually)

4.   *Catalog* (annually, with quarterly supplements)

5.   *Memento* (annually)

ISO is affiliated with the American National Standards Institute (ANSI).

Organisation Internationale de Normalisation (ISO)      *To order ISO documents in the U.S. contact:*
1, Rue de Varembe
Case Postale 56                                          American National Standards Institute
CH-1211 Geneva 20                                        ISO TC 97/SC6 Secretariat
Switzerland                                              1430 Broadway
41 22 341240 (phone); 41 22 333430 (fax)                 New York, NY 10018
Telex: 23 887 ISO CH                                     212/354-3300
Lawrence D. Eicher, Secretary General

### B.1.7.    National Information Standards Organization (NISO)

NISO was founded in 1939. NISO operates with a staff of three and budget of $200,000 with 66 members. NISO members are library and library networks, library associations, abstracting and indexing services, Federal agencies, publishers, and commercial, professional and educational organizations. NISO develops standards and promotes the voluntary use of technical standards in libraries and publishing and information services. NISO serves as a U.S. technical advisory group to TC 46 of the International Organization for Standardization (ISO). NISO maintains 25 standards committees. NISO is affiliated with the American National Standards Institute (ANSI). NISO was formerly (1984) the American National Standards Committee–Z39 and (1988) the National Information Standards Organization–Z39. NISO publishes the following material:

1.  *Information Standards Quarterly*, also cited as *The Voice of Z39: News about Library, Information Science, and Publishing Standards* (quarterly, $40 per year for nonmembers)

2.  *NISO Directory* (annually)

3.  List of Z39 published standards

National Information Standards Organization
P.O. Box 1056
Bethesda, MD 20817
301/975-2814
Patricia R. Harris, Executive Director

### B.1.8.    National Institute of Standards and Technology (NIST)

The National Bureau of Standards (NBS) name was changed in March of 1989 to National Institute of Standards and Technology (NIST). NBS was created by an act of Congress in 1901 to be the nation's measurement laboratory in the physical and engineering sciences. NIST is not a regulatory agency, but rather a laboratory used by industry, academia and Government alike as an independent source of technical information and advice. Its purpose is to help ensure the compatibility of measurement standards needed.

NIST's programs are directed towards reducing or removing technical barriers that impeded the prompt introduction or exploitation of new technologies. NIST is divided into the following laboratories:

1.  National Measurement Laboratory
2.  National Engineering Laboratory
3.  Institute for Computer Sciences and Technology
4.  Institute for Materials, Science and Engineering

*For general information:*
NIST Inquiry Service
Administration Building, Room E07
Gaithersburg, MD 20899
301/921-2318; 301/975-2762; 301/975-2786 (library)
Norma Redstone, Chief

*For publication information:*
NIST Technical Information and Publications
Administration Building, Room A537
Gaithersburg, MD 20899
301/921-2493
Donald Baker, Chief

*Mailing address:*   NIST
                     Washington, DC 20234

### B.1.8.1.  Federal Agencies

#### Department of Defense (DoD) Computer-Aided Acquisition and Logistics Support (CALS) Initiative

CALS is a joint DoD-industry effort to enable increased productivity through the integration of technical informa-
tion for a variety of critical functions spanning the range of weapon system acquisition, design, manufacturing, and
support. CALS will enable investments in automation to streamline the acquisition, design, manufacturing, and
operational support by elimination of redundant tasks and data. Wherever possible, DoD is adopting standards
from the private sector rather than creating unique DoD formats and protocols to meet its objectives. CALS stan-
dards can be expressed within the framework of the CALS core set, which includes the following standards:

1.  Functional Standards—tailored versions of MIL-STDs and specifications and contractual statements of
    work.

2.  Technical Standards—relevant conventions for the management, formatting, and physical or telecommuni-
    cations exchange of digitized text, graphics, alphanumerics, and other forms of data.

3.  Data Standards—data dictionaries and other conventions needed for access to databases.

The Connectivity Center at Informat in Dallas, Texas, began operations in January 1990. This not-for-profit organi-
zation will be serving as a resource and information center for the practical application of CALS standards.

CALS Policy Office
Department of Defense
The Pentagon, Room 2B322
Washington, DC 20301-8000
202/697-0051
Dr. Michael F. McGrath, Director

CALS Office, National Institute of Standards and Technology
Automated Production Technology Division
Building 233, Room B106
Gaithersburg, MD 20899
301/975-6641
David S. Bettwy, Deputy Division Chief

*For information about the CALS Connectivity Center, contact:*

Dr. John Wolcott
Texas Research and Technology Foundation
14785 Omicron Drive
San Antonio, Texas 78245
512/677-6000

Larry Samartin
Informat
1950 Stemmons Freeway, Suite 6038
Dallas, Texas 75207
214/746-5338

## B.2.    ESTABLISHED STANDARDS BY ORGANIZATION

### B.2.1.    Association for Information and Image Management (AIIM)

AIIM is currently active in the development of standards for electronic document image management. AIIM developed its first nonmicrographics standard (M44) in 1988 and is in the process of developing a number of additional standards. AIIM also has a number of project groups working on drafts for yet further standards.

#### B.2.1.1.    AIIM Adopted Standards

| AIIM NAME | TITLE/DESCRIPTION |
|---|---|
| MS44 (1988) | Recommended Practice for Quality Control of Image Scanners |
| X.440 | Scanner Quality Control Test Target |
| TR14 | Tiled Raster Interchange Format |
| TR17 | Facsimile and Its Role in Electronic Imaging |
| TR19 | Display Formats and Characteristics for Imagery |

#### B.2.1.2.    AIIM Standards Under Consideration or Revisions

| AIIM NAME | TITLE/DESCRIPTION |
|---|---|
| MS49 | Recommended Practice for Scanning from Roll Film and Microfiche |
| TR15 | Preparation of Documents for Image Capture Systems |

#### B.2.1.3.    AIIM Project Groups Developing Standards

| PROJECT | TITLE/DESCRIPTION |
|---|---|
| 80 | Content of Product Specification Sheets for Optical-Based Image Management Equipment and Software |
| 81 | Use of Electronic Image Printers in Electronic Document Imaging Systems |
| 82 | Optical Recording Media |
| 85 | Capture and Image Tagging |
| 86 | Enterprise Image Networking |
| 87 | Expunging of Optical Media |
| 76 | Optical Disk Package Labeling |
| 79 | Optical/Digital Imaging for Public Records |

### B.2.2.    American National Standards Institute (ANSI)

ANSI does not generally develop standards but primarily adopts standards that have been developed by other organizations. ANSI also serves as a reseller of published standards.

#### B.2.2.1.    Selected ANSI Adopted Standards

| ANSI NAME | TITLE/DESCRIPTION |
|---|---|
| X3.1 | Data Transmission Synchronous Signaling Rates |
| X3.4 | ASCII code standard defining the coded character set used extensively in information processing and communication |
| X3.15 | Specifies bit sequencing of the ASCII for serial by bit and serial by character transmissions |
| X3.16 | Describes character structure and parity for serial data transmission |
| X3.28 | Link control standard using a character oriented protocol |
| X3.44 | Describes system elements which determine information channel performance |
| X3.57 | Structure for Formatting Message Headings for Information Interchange Using the American National Standard Code for Information Interchange format Data Communication System Control |
| X3.66 | Advanced Data Communication Control Procedure (ADCCP) for bit-oriented link control |

| | |
|---|---|
| X3.100 | Interface Between Data Terminal Equipment and Data Circuit-terminating Equipment for Packet Mode Operation with Packet Switched Data Communications Networks |
| X3.122 | Computer Graphics Metafile for the Storage and Transfer of Picture Description Information |
| X3.129 | Information Systems Intelligent Peripherals Interface—Physical Level |
| X3.131 | Small Computer System Interface (SCSI); another high-level interface that allows all rotating memory devices to be addressed as logical devices |
| X3.139 | Fiber distributed data interface (FDDI) |
| IPI | Intelligent Peripherals Interface (IPI); a high-level interface that allows all rotating memory devices to be addressed as logical devices |
| X3T9 | Fiber Distributed Data Interface (FDDI); standard for the creation of a LAN using fiber-optic technology |

**B.2.3.    International Telecommunications Union (ITU)/Consultative Committee on International Telegraphy and Telephony (CCITT)**

*Comite Consultatif International Telegraphique et Telephonique* (Consultative Committee on International Telegraphy and Telephony) is one of four permanent committees of the International Telecommunications Union (ITU). CCITT adopts and maintains hundreds of standards in diverse areas, such as communications, definitions, symbols, testing, measuring, equipment connectivity, performance, software engineering, and conservation. CCITT has not developed standards that are specifically directed towards Digital Document Imaging Systems as a whole, but rather has developed standards for individual components and subcomponents. Only CCITT standards that directly relate to DDIS components have been included below.

*B.2.3.1.    Selected CCITT Adopted Standards*

| CCITT NAME | TITLE/DESCRIPTION |
|---|---|
| V.10 | Defines unbalanced interfaces between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE); same as RS-423 |
| V.11 | Defines the electrical characteristics of a balanced interface between DTE and DCE. (Same as RS-422) |
| V.24 | Defines the functions of all DTE/DCE interface lines |
| V.28 | Defines the electrical characteristics compatible with RS-232C |
| V.35 | Data Transmission at 48 kilobits per second (Kbps) using 60–108 kiloHertz (kHz) group-band circuits |
| X.20 | Public data network interface recommendations covering asynchronous, character-oriented equipment |
| X.21 | Public data network interface recommendations covering synchronous equipment |
| X.25 | Definition for public data network describing interface, link control, and packet exchange procedures |
| X.31 | Interconnection of X.25 and Integrated Services Digital Network (ISDN) |
| X.75 | Interface between public packet-switched data |
| X.121 | International Numbering Plan for Public Data Networks |
| X.214 | Transport Service Definition for Open Systems Interconnection for CCITT Applications |
| X.224 | Transport Protocol Specification for Open Systems Interconnection for CCITT Applications |
| X.215 | Session Service Definition for Open Systems Interconnection for CCITT Applications |
| X.225 | Session Protocol Specification for Open Systems Interconnection for CCITT Applications |
| X.400 | Protocol standards for the exchange of electronic mail (E-Mail) |
| X.401 | Message Handling Systems; Basic Service Elements and Optical User Facilities |
| X.408 | Message Handling Systems; Encoded Information Type Conversion Rules |
| X.409 | Message Handling Systems; Presentation Transfer Syntax and Notation |
| X.410 | Message Handling Systems; Remote Operations and Reliable Transfer Server |
| X.411 | Message Handling Systems; Message Transfer Layer |
| X.420 | Message Handling Systems; Interpersonal Messaging User Agent Layer |

| | |
|---|---|
| X.430 | Message Handling Systems; Access Protocol for Teletex Terminals |
| X.500 | Another networking standard that is currently under development |
| Group 3 | Modified Read Code (MRC); a raster compression scheme for one- and two-dimensional coding |
| Group 4 | Modified MRC (MMRC); a raster compression scheme for total two-dimensional coding |

### B.2.4.    Electronics Industries Association (EIA)

EIA performs many similar functions as those of the Institute of Electrical and Electronics Engineers (IEEE). EIA is active in the development of standards. The standards developed by EIA are directed towards the component and subcomponents of a Digital Document Imaging System.

#### *B.2.4.1.  EIA Adopted Standards*

| EIA NAME | TITLE/DESCRIPTION |
|---|---|
| RS-232C | Defines the interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) |
| RS-334 | Defines signal quality and timing requirements across the DTE/DCE RS-232C interface |
| RS-422 | Designed to replace RS-232C; defines electrical characteristics of balanced interfaces |
| RS-423 | Designed to replace RS-232C; defines electrical characteristics of unbalanced interfaces |
| RS-449 | Designed to replace RS-232C; defines mechanical and functional requirements |

### B.2.5.    Institute of Electrical and Electronics Engineers (IEEE)

IEEE is very active in the development and maintenance of standards. IEEE has adopted and maintains hundreds of standards in many areas, such as broadcasting, antenna, circuits, communications, computer science, cement, electromagnetics, glass, instrumentation, and microwave technology. Many of these standards have direct bearing on the design of a Digital Document Imaging System, although IEEE has developed no DDIS-specific standards.

#### *B.2.5.1.  IEEE Adopted Standards*

| IEEE NAME | TITLE/DESCRIPTION |
|---|---|
| 802.1 | Coordinates the interface between the two lowest levels of the Open Systems Interconnection (OSI) model with the top five levels |
| 802.2 | Divides the OSI level two Data link layer into two parts, called the Media Access Control (MAC) subtler and the Logical Link Control (LLC) subtler |
| 802.3 | Standard for Local Area Networks: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) |
| 802.4 | Standard for Local Area Networks: Token-Passing Bus Access Method and Physical Layer Specification |
| 802.5 | Standard for Local Area Networks: Token-Ring Access Method and Physical Layer Specification |
| 1058.1 | Software Project Management Plans |
| 1063 | Standard for Software User Documentation |

### B.2.6.    International Organization for Standardization (ISO)

ISO is very active in the development of standards. ISO's standards deal with the individual components and subcomponents of Digital Information Systems. Like ANSI, ISO adopts (or slightly modifies and adopts) another organization's published standard. However, ISO differs from ANSI in that it also initiates and develops standards.

#### *B.2.6.1.  Selected ISO Adopted Standards*

| ISO NAME | TITLE/DESCRIPTION |
|---|---|
| 1745 | Character-oriented link control; Similar to ANSI X3.28 |
| 3309 | High-level Data Link Control (HDLC); bit-oriented protocol that defines the frame structure |

| | |
|---|---|
| 4335 | HDLC; bit-oriented protocol that defines the protocol elements |
| 7478 | Data Communication Multilink Procedures |
| 7498 | Describes the reference model for Open Systems Interconnection (OSI); is presently being used as a framework for network architectures |
| 8072 | Open Systems Interconnection Transport Service Definition |
| 8073 | Open Systems Interconnection Connection Oriented Transport Protocol Specification |
| 8326 | Information Processing Systems Open Systems Interconnection Basic Connection Oriented Session Service Definition |
| 8327 | Information Processing Systems Open Systems Interconnection Basic Connection Oriented Session Protocol Specification |
| 8348 | Data Communications Network Service Definition |
| 8473 | Data Communications Protocol for Providing the Connectionless Mode Network Service |
| 8571 | File Transfer, Access, and Method (FTAM) |
| 8602 | Open Systems Interconnection Protocol for Providing the Connectionless Mode Transport Service |
| 8613 | Office Document Architecture; multipart standard on ODA |
| 8648 | Open Systems Interconnection Internal Organization of the Network Layer |
| 8649 | Open Systems Interconnection Service Definition for the Association Control Service Element |
| 8650 | Open Systems Interconnection Protocol Specification for the Association Control Service Element |
| 8802 | See IEEE 802.3 |
| 8822 | Open Systems Interconnection Connection Oriented Presentation Service Definition |
| 8823 | Open Systems Interconnection Connection Oriented Presentation Protocol Specification |
| 8831 | Open Systems Interconnection Job Transfer and Manipulation Concepts and Services |
| 8832 | Open Systems Interconnection Specification of the Basic Class Protocol for Job Transfer and Manipulation |
| 8878 | Data Communications Use of X.25 to Provide the OSI Connection-Mode Network Service |
| 8879 | Standard Generalized Markup Language (SGML); physically describes the composition of a document |
| 9069 | SGML Support Facilities—SGML Document Interchange Format (SDIF) |
| 9160 | Data Encipherment Physical Layer Interoperability Requirements |
| 9314 | Fiber Distributed Data Interface (FDDI) |
| 9316 | Small Computer System Interface (SCSI) |
| 9542 | Telecommunications and Information Exchange Between Systems End System to Intermediate System Routing Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless Mode Network Service (ISO 8473) |
| 9660 | Volume and File Structure of CD-ROM for Information Interchange |
| 10149 | Data Interchange on Read-Only 120 Millimeter Optical Data Disks (CD-ROM) First Edition |
| TR10029 | Information Technology Telecommunications and Information Exchange Between Systems Operation of an X.25 Interworking Unit |

## B.2.7.    National Information Standards Organization (NISO)

NISO is active in the development of standards for libraries.

### B.2.7.1.    NISO Adopted Standards

**NISO NAME  TITLE/DESCRIPTION**
Z39.58          Common Command Language for Online Interactive Information Retrieval

## B.2.8.    Federal Standards and Regulations

This section lists the Federal agencies currently active in the development of standards for DDIS. Generally speaking, Federal agencies will adopt a private organization's published standard either as is or with some modifications.

Federal employees often serve as committee members in the development of standards by these private organizations.

### B.2.8.1.  National Institute of Standards and Technology (NIST)/Federal Information Processing Standards (FIPS)

NIST is the Federal agency charged with developing standards for all other Federal agencies. NIST produces Federal Information Processing Standards Publications (FIPS PUBS). In addition to NIST, FIPS publications (FIPS PUBS) can be purchased from several organizations, for example, the National Technical Information Service (NTIS). FIPS PUBS are developed by the Institute for Computer Sciences and Technology and issued under the provisions of the Federal Property and Administrative Services Act of 1949. FIPS PUBS include standards, guidelines, tests, and program information documents. Almost all FIPS PUBS are equivalent, or similar, to ANSI, IEEE, and ISO standards.

### B.2.8.2.  Adopted FIPS PUBS

See *Appendix C.1* on page 57 for DDIS-related FIPS PUBS.

### B.2.8.3.  Military—Department of Defense (DoD)

The U.S. military (MIL), particularly the Department of Defense (DoD), has developed standards in diverse areas such as weapons, aviation, navigation, tracking, communications, and electronics. In general, the military specification (MIL SPEC) series of standards have been adopted from a direct equivalent or similar standard as developed by one of the private standard development organizations.

#### Computer-aided Acquisition and Logistics Support (CALS) Initiative

The DoD has instituted a new initiative for weapon system procurements. The implementation of CALS in the 1990s will be significant for engineering documentation. The Army is working CALS directives into its Digital Storage and Retrieval of Engineering Data System (DSREDS). The Air Force is working CALS directives into its Engineering Data Computer-Assisted Retrieval System (EDCARS). The Navy is working CALS directives into its Engineering Data Management Information and Control System (EDMICS) and Navy Automated Information Logistics System (NAILS) projects.

The following list details the major standards that have been adopted as a part of the CALS initiative and that are directly related to a Digital Document Imaging System.

| MIL SPEC NAME | TITLE/DESCRIPTION |
|---|---|
| MIL-STD-1840A | Automated Interchange of Technical Information |
| MIL-STD-28000 | Digital Representation for Communication of Product Data: Application Subsets—Initial Graphics Exchange Specification (IGES) |
| MIL-STD-8001 | Standard Generalized Markup Language (SGML) |
| MIL-D-CGM | Digital Representation for Communication of Illustration Data: CGM Application Profile |
| MIL-R-RASTER | Requirements for Raster Graphics Representation in Binary Format |
| FED-STD-1064 | Telecommunications: General Aspects of Group 4 Facsimile |
| FED-STD-1065 | Telecommunications: Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus |
| MIL-STD-1777 | Internet Protocol (IP) |
| MIL-STD-1778 | Transmission Control Protocol (TCP) |
| MIL-STD-1780 | File Transfer Protocol (FTP) |
| FIPS PUB 100 | Interface Between DTE and DCE for operation with Packet-Switched Data Networks |
| (FED-STD-1041) | (PSDN), or between two DTE's by Dedicated Circuit [Defense Data Network (DDN) X.25] |
| FIPS PUB 128 | Computer Graphics Metafile (CGM) (ANSI X3.122) |

**B.2.9.    Industry and *De Facto* Standards**

A *de facto* standard is a product or process that is so common that it is accepted as the norm. Generally speaking, industry always drives standards. Specifically speaking, it is in a corporation's best interest for their product to become the industry standard as this always will increase their sales and profit. Industry and standards development organizations are always at odds when it comes to the development of standards. This is because these organizations will not adopt a company's product or process as a standard unless the vendor forsakes its proprietary technology. If the company will not release it proprietary technology, a standards organization will often times still develop a standard by modeling the standard after the product or process of a particular company. Thus, the company may not always conform to the standard once it becomes published.

This following sections provide information on proprietary products or processes that may be considered as imaging system *de facto* standards.

### *B.2.9.1.    International Business Machines (IBM)*

#### Personal Computer Disk Operating System (PC DOS)

Developed for the IBM PC by Microsoft Corporation, DOS is the operating system standard for PC-based applications. Microsoft produces its own version of DOS (known as MS-DOS) for manufacturers of PC compatibles.

#### Token-Ring Local Area Network (LAN)

IBM's Token-Ring LAN is a baseband ring network that uses the IEEE 802.5 token-passing access method. IBM's Token-Ring LAN has a star-wired ring topology with a token-passing access method. This approach is emerging as the most popular LAN for PC-oriented implementations. Its chief benefit lies in a large-data and high-use environment since it does no suffer from the collision and retransmission problems that the Ethernet LAN has.

#### Binary Synchronous Communication (BSC)

A character-oriented protocol that performs supervisory functions in establishing, maintaining, and releasing data links, and error and flow control.

#### Synchronous Data Link Control (SDLC)

A bit-oriented protocol that performs supervisory functions in establishing, maintaining, and releasing data links, and error and flow control.

#### Systems Network Architecture (SNA)

SNA is a mainframe-based communications network whose goal is to provide a general framework for networking and distributed processing. SNA does not completely follow the International Organization for Standardization (ISO) Open Systems Interconnection (OSI) standard.

#### Network Basic Input/Output System (NETBIOS)

NETBIOS was introduced by IBM in 1984 as a software interface between a network adapter and an operating system. It offers an alternative to the high level structure established by the IEEE 802 committees and has emerged as a *de facto* standard to supply network transport services for PC-only LANs.

#### Mixed Object—Document Content Architecture (MO:DCA)

MO:DCA is a deliberate attempt to establish an image standard within the larger context of a unified document structure. MO:DCA identifies various function sets that classify a document or object as containing presentation data. Some of MO:DCA's goals are to provide better document sharing and furnish device independence.

## Video Graphics Array (VGA)

VGA is the most widely accepted high-resolution standard in color graphic display modes. Originally built into the system board of the IBM Personal System/2 (PS/2) Micro Channel computers, VGA was quickly adopted by the industry as the current standard in color displays for PC compatibles with industry standard buses.

### B.2.9.2.   Digital Equipment Corporation (DEC)

## Digital Data Communications Message Protocol (DDCMP)

A byte-count oriented protocol that performs supervisory functions in the govern establishing, maintaining, and releasing data links, and error and flow control.

## Digital Equipment Corporation Network (DECnet)

DECnet Local Area Network (LAN) is a family of software and hardware products based on the framework of Digital Network Architecture (DNA) for system interconnection. DNA is an OSI-like architectural model within which DEC designs and develops all of its communications products. Ethernet is one of the protocols used.

## Digital Network Architecture (DNA)

DNA is a minicomputer- based communications network whose goal is to provide a general framework for networking and distributed processing. DNA does not completely follow the International Organization for Standardization (ISO) Open Systems Interconnection (OSI) standard.

### B.2.9.3.   Xerox

## Ethernet

The Xerox Ethernet system is a baseband, contention network that uses CSMA/CD for access. Ethernet was created in 1976 as a joint effort between Xerox, Intel, and Digital Equipment. The IEEE 802.3 standard was derived from Ethernet and closely parallels it, but is not identical in all respects.

### B.2.9.4.   American Telephone & Telegraph (AT&T)

## StarLAN

AT&T's StarLAN Network is a low-cost CSMA/CD "personal computer network" that operates over standard unshielded, twisted-pair telephone wire. It encompasses each of the seven layers of the OSI reference model. At the physical and data link protocol levels, the AT&T StarLAN Network is the proposed IEEE 802.3 draft standard and its key parameters are consistent with 802.3 Ethernet standards.

### B.2.9.5.   General Motors (GM)

## Manufacturing Automation Protocol (MAP)

MAP is an evolving, user-defined LAN communication protocol standard for an automated factory. Its scope has been expanded to include other manufacturers (beyond GM) so as to launch an industry-wide specification. The goal of MAP is to create a factory communication network utility to which computer-based equipment can be easily attached. Standards are taken from IEEE, ISO, and EIA.

## Technical and Office Protocol (TOP)

TOP is closely linked to the activities surrounding the development of the MAP specification. TOP is trying to do for the technical and office communications networks what MAP is accomplishing for the factory. TOP uses FTAM just as MAP does.

### B.2.9.6.  Telephone Companies

In the United States, the telephone system has established phone line interface standards. Private Branch Exchange (PBX) is now being suggested by many manufacturers as an alternative to LANs. All modems intended for use in the telephone System must meet the telephone regulations for service compatibility as well as those regulations imposed by the Federal Communications Commission (FCC).

### B.2.9.7.  Datapoint

#### Attached Resource Computer Network (ARCnet)

The ARCnet Local Area Network (LAN) was developed in 1977. ARCnet is an example of a token-passing bus with a hybrid bus/star topology. ARCnet is similar to the IBM Token Ring LAN as it uses the token-passing protocol.

### B.2.9.8.  Digital Communications Associates (DCA)

#### IRMA Card

The IRMA card is a plug-in PC card that allows a PC to double as a mainframe 3270 terminal. The 3270 is a generic name for an IBM system component that communicates with a mainframe using IBM's Systems Network Architecture (SNA).

### B.2.9.9.  Aldus Corporation

#### Tagged Image File Format (TIFF)

TIFF was originally designed for desktop publishing applications. TIFF has emerged as a file format standard for scanned images. TIFF has recently undergone its fifth revision. TIFF handles the following four data classes:

1. B, for bilevel (1-bit)
2. G, for grayscale images
3. P, for palette color images
4. R, for RGB (red-green-blue) color images

### B.2.9.10.  ZSoft Corporation

#### PC Paintbrush (PCX)

PCX is a bit-mapped image file format created by ZSoft for PC Paintbrush, its graphics drawing program, and Frieze, its screen-grabbing utility.

### B.2.9.11.  Kofax Image Products

#### Kofax Compression/Decompression Boards

In PC, mini and mainframe computer systems, the Kofax board is the generally accepted board to provide both compression and decompression of raster images. Generally speaking (there is no standard), the image is compressed immediately after being scanned—before it is sent over the network to be stored onto optical disk. Each display device/workstation must also have a decompression board in order to display the image.

### B.2.9.12. Hewlett-Packard (HP) Corporation

#### Printer Command Language (PCL)

A printer requires a Page Description Language (PDL) to determine what type of graphic file and fonts to use for the print. A common printer used in imaging systems is the Hewlett-Packard LaserJet laser printer. This printer operates with Hewlett-Packard's Printer Command Language (PCL).

### B.2.9.13. Adobe Corporation

#### PostScript

PostScript first appeared in the Apple LaserWriter laser printer and was later available as an add-on software product to Hewlett-Packard's Printer Command Language (PCL), which is used by the Hewlett-Packard LaserJet and compatible laser printers. PostScript can be considered a *de facto* standard because of its prevalence in commercial arts, publishing, and printing industries. Most graphics printers incorporate the PostScript language.

# C. CROSS-REFERENCE OF CURRENT STANDARDS

Of the various groups of existing standards, two have particular significance to Digital Document Imaging Systems in Federal agencies. The first group is the standards published by the National Institute of Standards and Technology (NIST) in the Federal Information Processing Standards Publications (FIPS PUBS). The second group is the Government Open Systems Interconnection Profile (GOSIP). Every attempt has been made to ensure that the information contained in this section is current and correct. By their very nature, however, standards are dynamic. Not only are new standards being adopted, but also existing ones undergo revision or are augmented.

## C.1. FIPS PUBS CROSS-REFERENCE MATRIX

Table 1 below cross-references the FIPS PUBS issued by NIST with the standards promoted by the American National Standards Institute (ANSI), the Consultative Committee on International Telegraphy and Telephony (CCITT) of the International Telecommunications Union (ITU), the International Organization for Standardization (ISO), and other relevant organizations.

**Table 1.** *FIPS PUBS Cross-Reference Matrix*

| FIPS | ANSI | CCITT | ISO | OTHER | FIPS PUB NAME |
|------|------|-------|-----|-------|---------------|
| 1-2 | X3.4 X3.32 X3.41 | | | | Code for Information Interchange, its Representations, Subsets, and Extensions |
| 16-1 | X3.15 | | | | Bit Sequencing of the Code for Information Interchange in Serial-by-bit Data Transmission |
| 17-1 | X3.16 | | | | Character Structure and Character Parity Sense for Serial-by-bit Communication in the Code for Information Interchange |
| 18-1 | X3.25 | | | | Character Structure and Character Parity Sense for Parallel-by-bit Communication in the Code for Information Interchange |
| 31 | | | | | Guidelines for ADP Physical Security and Risk Management |
| 32-1 | X3.2 X3.17 X3.49 | | 1004 1073/1 | | Optical Character Sets for OCR |
| 40 | | | | | Guideline for Optical Character Recognition Forms |
| 42-1 | | | | | Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment |
| 56 | | | | | Guideline for Managing Multivendor Plug-Compatible ADP Systems |
| 60-2 | X.3T9/600 | | | | I/O Channel Interface |

| FIPS | ANSI | CCITT | ISO | OTHER | FIPS PUB NAME |
|---|---|---|---|---|---|
| 61-1 | X.3T9/666 | | | | Channel Level Power Control Interface (used in conjunction with 60-2) |
| 62 | | | | | Operational Specifications for magnetic tape subsystems (used in conjunction with 60-2) |
| 63-1 | X.3T9/780 | | | | Operational Specifications for variable block rotating mass storage subsystems (used in conjunction with 60-2) |
| 75 | | | | | Guideline on Constructing Benchmarks for ADP System Acquisitions |
| 76 | | | | | Guideline for Planning and Using a Data Dictionary System |
| 77 | | | | | Guideline for Planning and Management of Database Applications |
| 85 | X3.86 | | | | Optical Character Recognition (OCR) Inks |
| 86 | X3.64 | . | . | | Additional Controls for Use with American National Standard for Information Interchange |
| 89 | X3.93 | | | | Optical Character Recognition (OCR) Character Positioning |
| 90 | X3.99 | | | | Guideline for Optical Character Recognition (OCR) Print Quality |
| 94 | | | | | Guideline on Electrical Power for ADP Installations |
| 97 | | | | | Operational Specifications for fixed-block, rotating mass-storage subsystems |
| 100 | | X.25 | 8208 | | Interface Between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for Operation with Packet-Switched Data Communications Networks |
| 107 | | | | IEEE 802.2 IEEE 802.3 IEEE 802.5 | Local Area Networks (LAN): Baseband Carrier Sense multiple access with collision detection access method and physical layer specifications and link layer protocol |
| 111 | X3.91M | | | | Storage Module Interfaces (an alternative to 60-2) |

| FIPS | ANSI | CCITT | ISO | OTHER | FIPS PUB NAME |
|------|------|-------|-----|-------|---------------|
| 117 | | | 7487/3 | | 130mm (5.25-inch) Flexible Disk Cartridge Track Format Using Modified Frequency Modulation Recording at 7,958 bprad on Two Sides—1.9 tpmm (48 tpi) for Information Exchange |
| 118 | | | 7685 | | Flexible Disk Cartridge Labelling and File Structure for Information Exchange |
| 120 | X3.124 | | 7942 | | Graphical Kernel System (GKS) |
| 122 | | | | | Conformance Tests for FIPS PUB 100 Version of CCITT 1980 Recommendation X.25, etc. |
| 123 | | | 8211 | | Specification for a Data Descriptive File for Information Interchange (DDF) |
| 126 | X3.133 | | 8907 | | Database Language NDL |
| 127 | X3.135 | | 9075 | | Database Language Structured Query Language (SQL) |
| 128 | X3.122 | | | | |
| | .111 | | | | Optical Character Recognition (OCR)—Dot-Matrix Character Sets for OCR-MA |
| 130 | X3.129 X3.130 X3.132 X3.147 | | | | Intelligent Peripheral Interface (IPI) |
| 131 | X3.131 | | | | Small Computer System Interface (SCSI) |
| 133 | | v.22 v.26 v.26 *bis* | | | Coding and Modulation Requirements for 2,400-bps Modems |
| 134-1 | | v.27 *bis* v.27 *ter* v.32 | | | Coding and Modulation Requirements for 4,800-bps Modems |
| 135 | | v.29 | | | Coding and Modulation Requirements for Duplex 9,600-bps Modems |
| 136 | | v.22 | | | Telecommunications: Coding and Modulation Requirements for Duplex 600 and 1,200-bps Modems |
| 138 | | | | | Telecommunications: Electrical Characteristics of Balanced Voltage Digital Interface Circuits |

| FIPS | ANSI | CCITT | ISO | OTHER | FIPS PUB NAME |
|---|---|---|---|---|---|
| 139 | | | | | Interoperability and Security Requirements for Use of the Data Encryption Standard in the Physical Layer of Data Communication |
| 142 | | | | | Telecommunications: Electrical Characteristics of Unbalanced Voltage Digital Interface Circuits |
| 143 | | | | RS-449 | General Purpose 37-Position and 9-Position Interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) |
| 144 | X3.102 | | | | Data Communication Systems and Services User-Oriented Performance Parameters |
| 146 | | | | | Government Open Systems Interconnection Profile—GOSIP (see section C.2 on page 61 below) |
| 147 | | | | | Group 3 Facsimile Apparatus for Document Transmission |
| 148 | | | | RS-466 | Procedures for Document Facsimile Transmission |
| 149 | | | | EIA 536 | General Aspects of Group 4 Facsimile Apparatus |
| 150 | | | | EIA 538 | Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus |
| 151 | | | | IEEE 1003.1 | POSIX: Portable Operating System Interface for Computer Environments |
| 152 | | | 8879 | MIL M 28001 | Standard Generalized Markup Language (SGML) |
| 154 | | | | EIA 530 | High-Speed 25-Position Interface for Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) |
| 156 | X3.138 | | | | Information Resource Dictionary System (IRDS) |
| 157 | | | | AIIM MS44 | Guidelines for Quality Control of Image Scanners |
| 158 | | | | | X Window System |

## C.2.   GOSIP STANDARDS

*The Government Open Systems Interconnection Profile (GOSIP) encompass many standards developed by other organizations. The GOSIP standards are listed below by organization.*

*Federal Standards*

FIPS 107, Local Area Networks; Baseband Carrier Sense Multiple Access with Collision Detection Access Method and Physical Layer Specifications and Link Layer Protocol.

FIPS 100, Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Operation with Packet-Switched Data Communications Networks.

Federal Standard FED-STD 1041, Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Operation with Packet-Switched Data Communications Networks, National Communications System.

*Institute of Electrical and Electronics Engineers (IEEE)*

IEEE 754, Binary Floating Point Arithmetic.

*Electronics Industries Association (EIA)*

EIA-232C, Interface between Data Terminal Equipment and Data Communication Equipment Employing Serial Binary Data Interchange.

*International Organization for Standardization (ISO)*

Information Processing Systems—Open Systems Interconnection—Basic Reference Model, Ref. No. ISO 7498-1

Information Processing Systems—Data Communication—Use of X.25 to provide the OSI Connection Mode Network Service, IS 8878.

Information Processing Systems—Open Systems Interconnection—Network Service Definition, IS 8348.

Information Processing Systems—Open Systems Interconnection—Addendum to the Network Service Definition Covering Connectionless Data Transmission, IS 8348 AD1.

Information Processing Systems—Open Systems Interconnection—Addendum to the Network Service Definition Covering Network Layer Addressing, IS 8348 AD2.

Information Processing Systems—Open Systems Interconnection—Internal Organization of the Network Layer, DIS 8648.

Information Processing Systems—Open Systems Interconnection—Protocol for Providing the Connectionless Network Service, IS 8473.

Information Processing Systems—Open Systems Interconnection—Data Communication—X.25 Packet Level Protocol for Data Terminal Equipment (DTE), IS 8208.

7-bit Coded Character Set for Information Processing Interchange, ISO-646, 1973.

Information Interchange—Representation of Local Time Differentials, ISO-3307, 1975.

Information Processing Systems—Open Systems Interconnection—Working Draft—End System to Intermediate System Routing Exchange Protocol for use in Conjunction with ISO 8473.

Information Processing Systems—Open Systems Interconnection—Transport Service Definition, ISO IS 8072, 1984.

Information Processing Systems—Open Systems Interconnection—Transport Protocol Specification, ISO IS 8073, 1984.

Information Processing Systems—Open Systems Interconnection—Session Service Definition, ISO IS 8326, 1987.

Information Processing Systems—Open Systems Interconnection—Session Protocol Specification, ISO IS 8327.

Information Processing Systems—Open Systems Interconnection—File Transfer, Access and Management Part 1: General Introduction, ISO 8571-1.

Information Processing Systems—Open Systems Interconnection—File Transfer, Access and Management Part 2: The Virtual Filestore Definition, ISO 8571-2.

Information Processing Systems—Open Systems Interconnection—File Transfer, Access, and Management Part 3: File Service Definition, ISO 8571-3.

Information Processing Systems—Open Systems Interconnection—File Transfer, Access, and Management Part 4: File Protocol Specification, ISO 8571-4.

Information Processing Systems—Open Systems Interconnection—Connection-Oriented Presentation Service Definition, ISO 8822.

Information Processing Systems—Open Systems Interconnection—Connection-Oriented Presentation Protocol Specification, ISO 8823.

Information Processing Systems—Open Systems Interconnection—Service Definition for Association Control Service Element—Part 2: Association Control, IOS 8649.

Information Processing Systems—Open Systems Interconnection—Protocol Specification for Association Control Service Element: Association Control, ISO 8650.

Information Processing Systems—Open Systems Interconnection—Specification of Abstract Syntax Notation One (ASN.1), ISO 8824.

Information Processing Systems—Open Systems Interconnection—Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1), ISO 8825.

Information Processing Systems—Data Communications—High-Level Data Link Control Procedures—Description of the X.25 LAPB-compatible DTE Data Link Procedures, IS 7776.

Information Processing Systems—Data Interchange—Structures for the Identification of Organizations, IS 6523, 1984.

Information Processing Systems—Local Area Networks—Part 2: Logical link control, DIS 8802/2.

Information Processing Systems—Local Area Networks—Part 3: Carrier sense multiple access with collision detection, IS 8802/3.

Information Processing Systems—Local Area Networks—Part 4: Token-passing bus access method and physical layer specifications, IS 8802/4.

Information Processing Systems—Local Area Networks—Part 5: Token ring access method and physical layer specifications, IS 8802/5.

*International Telecommunications Union (ITU)/Consultative Committee on International Telegraphy and Telephony (CCITT)*

CCITT Recommendation X.25-1984, Interface Between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for Terminals Operating in the Packet Mode on Public Data Networks.

CCITT Recommendation, X.400, Message Handling Systems: System Model-Service Elements.

CCITT Recommendation X.401, Message Handling Systems: Basic Service Elements and Optical User Facilities.

CCITT Recommendation X.408, Message Handling Systems: Encoded Information Type Conversion Rules.

CCITT Recommendation X.409, Message Handling Systems: Presentation Transfer Syntax and Notation.

CCITT Recommendation X.410, Message Handling Systems: Remote Operations and Reliable Transfer Server.

CCITT Recommendation X.411, Message Handling Systems: Message Transfer Layer.

CCITT Recommendation X.420, Message Handling Systems: Interpersonal Messaging User Agent Layer.

CCITT Recommendation X.430, Message Handling Systems: Access Protocol for Teletex Terminals.

CCITT Recommendation X.214, Transport Service Definition for Open Systems Interconnection for CCITT Applications.

CCITT Recommendation X.224, Transport Protocol Specification for Open Systems Interconnection for CCITT Applications.

CCITT Recommendation X.215, Session Service Definition for Open Systems Interconnection for CCITT Applications.

CCITT Recommendation X.225, Session Protocol Specification for Open Systems Interconnection for CCITT Applications.

CCITT Recommendation X.121, International Numbering Plan for Public Data Networks.

CCITT Recommendation V.35, Data Transmission at 48 kilobits per second (Kbps) using 60–108 kHz group band circuits.

# D. ADDITIONAL RESOURCES

Digital imaging technology can be a challenging subject to learn. Additional resources are available for further information in the form of printed material, such as books and serials/journals. Conferences/training seminars and consultants/research groups provide other sources of valuable information. This appendix contains lists of these various resources. These lists are not intended to be inclusive, nor are they endorsements. They are provided solely as general references. Where appropriate, addresses and phone numbers are also given.

## D.1. PRINTED MATERIAL

Resources in the form of printed material include books and serials/journals. Many of the books listed below undergo periodic revision due to the rapidly changing technology covered.

### D.1.1. Books

*Converting Information for WORM Optical Storage: A Case Study Approach*, Meckler Publishing, 11 Ferry Lane West, Westport, CT 06880; 203/226-6967.

*Datapro Reports on Document Imaging Systems*, Datapro Research, McGraw-Hill Publishing Company, 1990; 800/328-2776.

*Digital Imaging Technology Assessment*, Digital Document Storage project, NASA Scientific and Technical Information (STI) Facility, August 1989.

*Document Image Processing for IBM Users*, Xephon Consultancy Report, Xephon Technology Transfer Limited, September 1989.

*Feasibility Report for Code 408 Technical Documentation Library (TDL) Document Imaging System (DIS)*, Dale E. Houghtaling, EER Systems Inc., 1989.

*Handbook of Optical Memory Systems Update, The*, C. Peter Waegemann, Optical Disk Institute, 1989.

*Image Storage and Retrieval Systems*, Marc D'Alleyrand, McGraw-Hill Publishing Company, 1989.

*Index and Directory of Industry Standards*, Information Handling Services Inc., 1989.

*Information Management Sourcebook: The AIIM Buying Guide and Membership Directory*, Association for Information and Image Management (AIIM), 1100 Wayne Avenue, Suite 1100, Silver Spring, MD 20910; 301/587-8208.

*Information Technology Standardization: Theory, Process, and Organization*, Carl F. Cargill, Digital Equipment Corporation, 1989.

*Survey of Optical Disk Technologies*, Govind G. Idnani, MITRE Corporation, 1988.

*Space Station Technical and Management Information System (TMIS): Technology Assessment of Optical Storage Devices for the TMIS Program*, Clarence Catoe, Boeing Computer Services, July 29, 1988.

**D.1.2.   Serials/Journals**

*Data Conversion Newsletter*, Image Publishing, P.O. Box 3149, Westport, CT 06880; 203/222-9310

*Image Technology Report*, Microfilm Publishing Inc., P.O. Box 950, Larchmont, NY 10538; 914/830-3044.

*IMC Journal*, International Management Congress (IMC), 345 Woodcliff Drive, Fairport, NY 14450; 716/383-8330.

*INFORM*, Association for Information and Image Management, 1100 Wayne Avenue, Suite 1100, Silver Spring, MD 20910; 301/587-8202.

*Micrographics Newsletter*, Microfilm Publishing Inc., P.O. Box 950, Larchmont, NY 10538; 914/830-3044

*Modern Office Technology*, Penton Publishing, 1100 Superior Avenue, Cleveland, OH 44114; 216/696-7000.

*Optical Information Systems Magazine*, Meckler Publishing, 11 Ferry Lane West, Westport, CT 06880; 203/226-6967.

*Optical Information Systems Update*, Meckler Publishing, 11 Ferry Lane West, Westport, CT 06880; 203/226-6967.

*Optical Memory News*, Rothchild Consultants, 256 Laguna Honda Boulevard, San Francisco, CA 94116-1496; 415/681-3700.

*TECHinfo*, Graphic Communications Association, 1730 North Lynn Street, Suite 604, Arlington, VA 22209-2085; 703/841-8160.

*Today's Office*, FM Business Publications Inc., 1225 Franklin Avenue, Garden City, NY 11530; 516/739-0335.

## D.2.   CONFERENCES/TRAINING SEMINARS

AIIM Show and Conference, Association for Information and Image Management, 1100 Wayne Avenue, Suite 1100, Silver Spring, MD 20910; 301/587-8202.

Federal Computer Conference, National Council for Education on Information Strategies, P.O. Box 13376, Silver Spring, MD 20911-3376.

Graphic Communications Association, 1730 North Lynn Street, Suite 604, Arlington, VA 22209-2085; 703/841-8160.

Kalthoff Group, The, P.O. Box 8931, Cincinnati, OH 45208; 513/871-6808.

Optical Information Systems Conference, Meckler, 11 Ferry Lane West, Westport, CT 06880; 203/226-6967.

Rothchild Consultants, 256 Laguna Honda Boulevard, San Francisco, CA 94116-1496; 415/681-3700.

U.S. Professional Development Institute, 1734 Elton Road, Suite 221, Silver Spring, MD 20903; 301/445-4400.

## D.3.   CONSULTANTS/RESEARCH GROUPS

Anderson Consulting, 69 West Washington Street, Chicago, IL 60602; 312/507-6433.

BIS CAP International, One Longwater Circle, Norwell, MA 02061; 617/982-9500.

Delphi Consulting Group, 50 Staniford Street, Boston, MA 02114; 617/723-7446.

Document Automation Consulting, 266 Alvarado Road, Berkeley, CA 94705; 415/486-1024.

International Data Corporation/Avante, 5 Speen Street, Framingham, MA 01701; 508/872-8200.

Kalthoff Group, The, P.O. Box 8931, Cincinnati, OH 45208, 513/871-6808.

Rothchild Consultants, 256 Laguna Honda Boulevard, San Francisco, CA 94116-1496; 415/681-3700.

Yankee Group, The, 200 Portland Street, Boston, MA 02114; 617/367-1000.

# E. GLOSSARY

This appendix contains a glossary relevant terms used and acronyms used in the body of this report. Every effort has been made it complete and accurate, but many terms are constantly being revised or added in the field of digital imaging technology. The references that were consulted in compiling this glossary are given in section E.2 on page 91.

## E.1. GLOSSARY OF TERMS

**access time:** The time required to gain entry to a specific location on a memory device.

**address:** A unique designation used to identify the location in storage of an item of data or the destination of a packet of data sent from one computer terminal to another.

**ADCCP:** Advanced Data Communication Control Procedure.

**ADP:** Automated Data Processing.

**ADPE:** ADP Equipment.

**AI:** Artificial intelligence; a subfield of computer science concerned with the concepts and methods of symbolic inference by a computer and the symbolic representation of the knowledge to be used in making inferences.

**AIIM:** Association for Information and Image Management; professional association for manufacturers, vendors, and professional users of information and image management equipment, products, and services.

**AIM:** NASA's Automated Information Management.

**algorithm:** A special mathematical procedure for solving a particular type of problem.

**American National Standards Institute:** See "ANSI."

**American Standard Code for Information Interchange:** See "ASCII."

**analog:** In communications, transmission employing variable and continuous waveforms to represent information values, where interpretation by the receiver is an approximation (quantization) of the encoded value. Compare "digital."

**ANSI:** American National Standards Institute; the principal U.S. group that is concerned with developing standards. ANSI is a nonprofit organization supported by Government, industry, and other professional organizations.

**antialiasing:** The algorithmic process of smoothing short horizontal and vertical lines made by the computer to represent a diagonal line in a digitized image. (The stair-stepped effect of the horizontal and vertical lines is known as aliasing.)

**APP:** Application Portability Profile; standard adopted by the National Institute of Standards and Technology that incorporates X Window System technology. Compare "X Window System."

**application software:** Programs that perform useful functions in the processing or manipulating of data for the user; includes payroll programs, word processors, text editors, spreadsheets, and other programs that enable the useful production of information.

**archival:** This term, when used for media means that it is readable (and sometimes writable) for five years or longer.

**ARPANET:** Advanced Research Projects Agency Network; a subset of the U.S. Department of Defense's Defense Data Network. Compare "DDN."

**artificial intelligence:** See "AI."

**ASCII:** American Standard Code for Information Interchange; a character set established by ANSI to represent alphanumerics in eight bits—seven plus parity—and used for asynchronous transmission.

**aspect ratio:** The relationship between the length and width of a two-dimensional area.

**Association for Information and Image Management:** See "AIIM."

**asynchronous transmission:** Transmission in which time intervals between transmitted characters might be of unequal length. Transmission is controlled by start and stop bits at the beginning and end of each character.

**average access time:** The average time a computer takes to deliver requested information, measured from the instant of request to the instant of delivery.

**Automated Data Processing:** See "ADP."

**b:** Abbreviation for "bit." See "bit."

**B:** Abbreviation for "byte." See "byte."

**backup:** A copy of stored data.

**bandwidth:** The range of frequencies that can be passed through a channel.

**batch processing:** A data processing technique in which input data is accumulated and processed in groups or batches.

**baud:** A measure of the speed of data transmission; equal to the number of units, usually bits, sent per second.

**Baudot:** Data transmission code in which five bits represent one character. Use of letters/figures shift enables 64 alphanumeric characters to be represented. Baudot is used in many teleprinter systems with one start bit and 1.42 stop bits added.

**BCD:** Binary Coded Decimal; a 6-bit code for structuring data. Compare "EBCDIC."

**Bernoulli box:** A storage medium for computer data that uses a removable cartridge containing a flexible disk (capacity ranges from 10MB to 44MB, depending on the model). A Bernoulli box combines the storage capacity and reliability of a hard-disk drive with the portability and durability of a diskette.

**Binary Coded Decimal:** See "BCD."

**binary digit:** See "bit."

**bit:** Short for "binary digit," a bit is a unit of computer information that can be represented by an electrical impulse, a magnetized spot, or a hole whose presence or absence indicates data.

**bit map:** Representation of image data where each pixel has a corresponding memory element. The bit map contains a bit for each point or dot on the screen, allowing for very fine resolution since any point on the screen can be addressed.

**bpi:** Bits per inch.

**bps:** Bits per second.

**broadband:** See "wideband."

**byte:** A group of adjacent bits, often shorter than a word, that a computer processes as a unit, for example, an 8-bit byte.

**CALS:** Computer-aided Acquisition and Logistics Support; a U.S. Department of Defense (DoD) initiative to provide a standardized method of exchanging text and images among multiple computer hardware and software platforms. CALS is actually a superset of existing standards governing formats for computer graphics files, document typesetting tags, and storing scanned images. See "CGM," "SGML," and "TIFF."

**capacity:** The capacity of a computer storage system is generally expressed in the number of bytes of data that the medium can hold. The capacity of a magnetic disk or diskette is affected by the density of bits per inch, and the number of tracks available for use.

**CAR:** Computer-Assisted Retrieval; a retrieval system that combines a file of records on microfilm with a computerized index.

**cartridge:** In optical technology, an enclosure, generally of plastic, in which an optical medium is kept for protection; also called a cassette.

**CAV:** Constant Angular Velocity; describes a disk that always spins at the same rotation rate. Compare "MCAV" and "ZCAV."

**CCD:** Charged-Coupled Device; a semiconductor device known as a recirculating memory, used frequently in memory chips for digital scanners because of the CCD's short-term memory and fast access capabilities.

**CCITT:** Consultative Committee on International Telegraphy and Telephony (*Comite Consultatif International Telegraphique et Telephonique*) is a permanent committee of the International Telecommunications Union (ITU). The CCITT sets standards to facilitate public telecommunications. Compare "ITU."

**CCS 7:** A network signaling standard for ISDN that incorporates information from databases in order to offer advanced network services.

**CD:** Compact Disc, the trademarked name for the laser-read digital audio disk, 12 centimeters in diameter, developed jointly by Philips and Sony; Compare "CD-ROM."

**CDP:** Compound Data Processing; the integrated processing of full text, voice, images, and digital representations of images.

**CD-ROM:** Compact Disc–Read-Only Memory; a version of the audio Compact Disc intended to store data that cannot be added to or changed. Compare "CD."

**central office:** The building in which common carriers terminate customer circuits (also known as central exchange).

**Central Processing Unit:** See "CPU."

**Centronics interface:** A Centronics Corporation 8-bit parallel interface for computer peripheral, such as printers, that is used by the computer industry as a standard. Compare "parallel interface" and "serial interface."

**CGM:** Computer Graphics Metafile; a standard file format for electronically stored images.

**character:** Letter, numeral, punctuation, control figure, or any other symbol contained in a message.

**character oriented:** A transmission procedure or communications protocol that includes control information encoded in fields of one or more bytes.

**character recognition:** The automated identification of letters, numbers, punctuation, or symbols. Compare "OCR."

**characters per second:** See "cps."

**CICS:** Customer Information Control System; an operating system for IBM mainframes.

**CIM:** Computer-Input Microfilm; microfilm containing data that can be automatically scanned and converted into a computer-usable form.

**clock:** Shorthand term for the source(s) of timing signals used in synchronous transmission. More generally, the source(s) of timing signals sequencing electronic events.

**cluster:** A collection of terminals or other devices in a single location.

**CLV:** Constant Linear Velocity; describes a disk that rotates more slowly when outer radii are being scanned. Compare "MCLV" and "QLV."

**COM:** Computer-Output Microfilm; dry-processed microfilm produced by converting computer-generated signals from digital data into an image.

**common carrier:** A private data communications utility company that furnishes communications services to the general public.

**Common User Access:** See "CUA."

**communications network:** Connected, geographically separated facilities for the rapid reception of, transmission of, and/or relaying of electrical impulses for reproduction as printed messages, pictures, or other data.

**communications protocol:** The rules governing the exchange of information between devices on a data line.

**Compact Disc:** See "CD."

**Compact Disc–Read-Only Memory:** See "CD-ROM."

compatibility: The ability to work together without the need for any modifications. May apply to computer hardware or software.

composite link: The line or circuit connecting a pair of multiplexors or concentrators; the circuit carrying multiplexed data.

Compound Data Processing: See "CDP."

compression: See "data compression."

compression algorithms: The formulae according to which digital data sets representing images are compressed.

Computer-aided Acquisition and Logistics Support: See "CALS."

Computer-Assisted Retrieval: See "CAR."

computer graphics: Graphs, charts, and pictures generated by means of a computer.

Computer Graphics Metafile: See "CGM."

Computer-Input Microfilm: See "CIM."

Computer-Output Microfilm: See "COM."

configuration: The selection and arrangement of the hardware used for a computerized system.

Constant Angular Velocity: See "CAV."

Consultative Committee on International Telegraphy and Telephony: See "CCITT."

conversion: The changing of data from one form of representation to another. Examples of conversion include large-scale, format changes required by changes to a computer system, changes needed when data is switched to a computer system that is not compatible, changes resulting from programming language or database management system changes, and automated changes from scanning a paper copy to produce an electronic or micrographic copy.

cps: characters per second; the duty cycle rate that is a measure of speed of either an impact printer or a transmission device.

CPU: Central Processing Unit; the portion of a computer that directs the sequence of operations and initiates the proper commands to the computer for execution.

CRT: Cathode-ray tube; used for television screens and computer video display terminals (VDT).

CUA: Common User Access; the portion of the Systems Application Architecture (SAA) standard developed by IBM that prescribes how information should be presented on the computer display and how function keys should work on the keyboard so that applications that comply with CUA will look and work in the same familiar way to users. For example, CUA prescribes the use of the F1 function key to access the help screen for an application. See "SAA."

Customer Information Control System: See "CICS."

**DACS:** Directory Access and Control System.

**DAT:** Digital Audio Tape.

**Data Circuit-terminating Equipment:** See "DCE."

**data compression:** An encoding technique that saves computer storage space by eliminating not only characters or data that are not needed, but also empty fields, and gaps.

**data decompression:** A decoding technique that reverses compression to convert what was compressed to the original data.

**data file:** See "file."

**data integrity:** A performance measure based on the rate of undetected errors.

**data rate:** The rate at which data moves through, out of, or into a device; frequently modified by peak, burst, instantaneous, sustained, average, or other indicators of the conditions of measurement.

**database:** An extensive and comprehensive set of records collected and organized in a meaningful manner to serve a particular purpose.

**database management system:** See "DBMS."

**DATAROM:** Sony's acronym for its 5.25-inch Optical Read-Only Memory (OROM) disk.

**DBMS:** Database management system; a computer method for organizing and manipulating a database.

**DCA:** Document Content Architecture.

**DCE:** Data Circuit-terminating Equipment; devices that provide the functions required to establish, maintain, and terminate a data transmission connection; for example, a modem.

**DDE:** Dynamic Data Exchange; method for linking and exchanging data between applications written for the Microsoft Windows environment.

**DDF:** Data Descriptive File.

**DDIS:** Digital Document Imaging System; A system designed to convert documents into binary (digital) code representing an image of the document and to store that code onto optical media. The digitized image can be reproduced as a paper copy of the original document or displayed as an image of the original document upon request.

**DDN:** Defense Data Network; a U.S. Department of Defense communications network that links a number of authorized computers at military installations, Government contractors, and universities.

**decompression:** See "data decompression."

*de facto:* in actuality or reality; from Latin meaning "according to the fact."

*de facto* **standard:** Used in the computer industry to describe a product, process, or format that is so widely accepted and used that it is really a standard even though it has not been officially sanctioned by a standards-issuing organization.

**defect:** An irregularity in a medium that disturbs its ability to store recorded data.

**DES:** Data Encryption Standard.

**device driver:** A software package that manages the interface between the CPU and a specific external device such as a tape or disk drive.

**DIA:** Document Interchange Architecture.

**DIF:** Document Interchange Facility.

**digital:** Referring to communications procedures, techniques, and equipment by which information is encoded as either a binary 1 or 0; the representation of information in discrete binary form, discontinuous in time. Compare "analog."

**digital data:** Information transmitted in a coded form (from a computer) represented by discrete signal elements.

**Digital Document Imaging System:** See "DDIS."

**digital paper:** A relatively new optical storage flexible media that has a storage capacity of one gigabyte on a single-sided 5.25-inch optical disk.

**digital scanner:** An optical device that converts graphics images into digital data that can be processed by a computer.

**digitizing:** The conversion of analog (that is, nondigital) data, such as graphics images, into digital data.

**disk:** A medium for randomly addressable data storage.

**Disk Operating System:** See "DOS."

**disk partitioning:** The division of one large disk area into smaller user-defined storage blocks.

**dithering:** process of arranging pixels (usually multiple bit) that contain grayscale information in to patters, called grains, that have various shapes to produce the illusion of large and small dots.

**DLL:** Dynamic Link Library.

**document:** Written, printed, or graphic information in paper form. Compare "external document" and "internal document."

**DoD:** U.S. Department of Defense.

**DOS:** Disk Operating System; the operating system for computers that manages the physical resources of a computer such as disks, memory, or displays. UNIX, MS-DOS, OS/2, and VMS are examples. See "OS."

**dots per inch:** See "dpi."

**downloading:** The process of sending configuration parameters, operating software, or related data from a central source to remote stations.

**dpi:** dots per inch; a measure of the resolution of the output of a printer.

**drive:** A machine for reading and, when possible, writing a data storage medium (disk, tape, card, or otherwise); can be optical, magnetic, etc.

**driver software:** See "device driver."

**DTE:** Data Terminal Equipment.

**dumb terminal:** Both hard-copy and VDT-type ASCII asychronous terminals that do not use a data transmission protocol and usually send data one character at a time.

**duty cycle:** measure of the capacity to perform a specific process per unit time. For example, a computer printer's duty cycle is quantified as the maximum number of pages per minute (ppm) that it can physically print.

**Dynamic Data Exchange:** See "DDE."

**dynamic thresholding:** the capability of a digital scanner to adjust automatically the sensitivity it has to reflected light to allow better contrast between light and dark areas of an image to be scanned.

**EBCDIC:** Extended Binary Coded Decimal Interchange Code; An 8-bit character code that provides for 256 different bit patterns. Used primarily in IBM mainframes. Compare "BCD."

**EDI:** Electronic Data Interchange.

**EIA:** Electronic Industries Association. Formerly the Radio Manufacturers Association (which later absorbed the Magnetic Recording Industry Association).

**embedded SQL:** embedded Structured Query Language (SQL); SQL is made a part of a compiled program's code with an assurance of portability. Compare "SQL."

**emulation:** The imitation of a computer system, performed by a combination of hardware and software, that allows programs to run between incompatible systems.

**encryption:** The scrambling or conversion of data, prior to transmission, to a "code" that masks the meaning of data to any unauthorized recipient and requires a key for decryption.

**erasable:** Rewritable. Compare "M-O."

**external document:** a collection of information (text, graphics images, or both) contained on paper, microform, or other nondigital media before it is converted into a digital form that can be processed by a computer. Compare "internal document" and "source document."

**Ethernet:** A Local Area Network standard developed by the Xerox Corporation for data communications.

**facsimile:** See "fax."

**fax:** (noun) The electronic transmission of a document page image by a fax machine; or the hard-copy output produced by a fax machine. (verb) To send document page images via a fax machine.

**fax board:** An add-in board that enables a PC workstation to send or receive a facsimile (fax) transmission. A received fax can be output on a printer connected to the PC. Compare "PC-to-fax interface."

**FCC:** Federal Communications Commission.

**FDDI:** Fiber-optic Distributed Data Interface; an ANSI standard for dual counter-rotating token rings with a bandwidth of 100 Mbps.

**FDX:** Full DupleX; simultaneous, two-way, independent transmission in both directions (4-wire).

**Federal Information Processing Standard:** See "FIPS."

**Federal Information Processing Standard Publication:** See "FIPS PUB."

**file:** A collection of records, containing the same field structure, and stored under a single name.

**file maintenance:** The activity of keeping the data in a file current by adding, changing, and deleting items.

**file organization:** Details concerning the programmed data file structure and how data in the file can be retrieved and viewed.

**file server:** A special-purpose computer where shared software resources are stored, including the network software that monitors network operation. The file server software manages access to a shared disk and the data on it. File server software is designed specifically for networking and is built to handle the sharing of files in a multiuser environment.

**FIPS:** Federal Information Processing Standard; a standard developed by the Institute for Computer Sciences and Technology under the direction of National Institute of Standards and Technology (NIST). Compare "NIST."

**FIPS PUB:** Federal Information Processing Standard Publication; a commercially available publication that contains a specific standard adopted by the National Institute of Standards and Technology (NIST). Compare "NIST."

**firmware:** Information or computer instructions stored in Read-Only Memory (ROM). Compare "software."

**font:** See "type font."

**frequency:** The number of cycles per second of a periodic phenomenon; unit is hertz (Hz); see "wave."

**FTAM:** File Transfer, Access, and Management.

**full-text database:** A collection of primary source records in their entirety, such as the complete text of successive issues of a newspaper, or court cases. See "database" and "text files."

**G:** Giga; one billion.

**gateway:** A device or program that performs a protocol conversion function between two dissimilar networks or pieces of equipment. Also used to describe the X.75 interface between two public data networks.

**Gb:** Gigabits; approximately one billion bits. Actually, $2^{30}$, or 1,073,741,824 bits.

**GB:** Gigabytes; approximately one billion bytes. Actually, $2^{30}$, or 1,073,741,824 bytes.

**GDDM:** Graphical Data Display Manager.

**GKS:** Graphical Kernel System.

**GOSIP:** Government Open Systems Interconnection Profile.

**Graphical User Interface:** See "GUI."

**graphics:** Artwork such as graphs, charts, and pictures.

**grayscale:** The variations between black and white that can be detected and passed to a computer by a scanner. This assists in the process of analyzing and then synthesizing a color image for monochrome display or printing. The number of levels of gray that are detectable depends upon the number of bits per pixel that the scanner records. grayscale levels in scanners are 16, 64, and 256.

**Group 3 facsimile:** A digital facsimile device CCITT classification relating to a speed of transmission of 9.6 Kbps or less (approximately one minute per page), over a public switched telephone network, the use of digital transmission, and a complex data compression capability.

**Group 4 facsimile:** A digital facsimile device CCITT classification relating to high-speed transmission of 56 Kbps (approximately 10 seconds per page), over a digital network, and a complex data compression capability. Group 4 machines have wide scanners that can handle large documents. Group 4 is further divided into 3 Classes each of which has a mandatory resolution standard.

**GUI:** Graphical User Interface; method of displaying information on a computer display in windows using icons to represent computer programs, functions, and peripherals. A GUI enables a user to manipulate the information in windows using a pointing device to access drop-down and pop-up menus, scroll bars, and radio buttons. In addition, the user can perform computer a operation by pointing to an icon without having to know the underlying computer commands or syntax that performs the operation.

**halftone:** A image reproduction technique that includes grays as well as black and white.

**hard-disk drive:** A rigid, high-capacity, magnetic storage medium, usually built into a computer, but now sometimes removable. See "Winchester disk."

**HDLC:** High-level Data Link Control; the communications protocol defined by the International Organization for Standardization (ISO).

**header:** The control information added to the beginning of a message transmission block or packet.

**hierarchy:** An arrangement of memory device types connected to form a series with increasing values of one parameter (such as accessibility) and decreasing values of another parameter (such as cost per bit).

**High-level Data Link Control:** See "HDLC."

**host computer:** The primary computer or the one in control in a network of computers.

**host interface:** The link between a host computer and a communications processor or network.

**Huffman code:** A code used for one-dimensional data compression in the CCITT Group 3 digital facsimile standard. Compare "modified Huffman code."

**hypermedia:** A linkage with typed commands, of different media, such as text, video, graphics, or audio material.

**hypertext:** A linkage of textual data both within and across documents.

**Hz:** Hertz; a unit of frequency equal to one cycle per second.

**IBM:** International Business Machines Corporation.

**ICR:** Intelligent Character Recognition; a type of Optical Character Recognition (OCR) that employs a combination of matrix-matching and feature-extraction algorithms to identify characters from a wider range of type styles and in more complex page layouts. Compare "OCR."

**IEEE:** Institute of Electrical and Electronics Engineers.

**IGES:** Initial Graphics Exchange Specification; an information model describing data elements and relationships used in Architecture/Engineering/Construction (AEC). Compare "PDES."

**IMG:** An image file format created by Digital Research Inc. for the Graphics Environment Manager (GEM). This format is used by Xerox's Ventura Publisher.

**image resolution:** The sharpness of an image such as that created by a scanner, a printer, or a CRT monitor. Image resolution may be expressed in dots per inch or lines per millimeter.

**imaging area:** The portion of a sheet of paper upon which an output device, such as a printer, can place a mark.

**import/export slot:** The slot used to add disks to or remove disks from an optical disk jukebox; also referred to as an exchange slot or mailbox.

**IMS:** Information management system; method of organizing the storage, maintenance, and retrieval of information; generally computerized and available in real time.

**index:** In computers, an ordered list of files that have been entered. In a publication, an alphabetical list of keywords or some other classification of information contained in that publication, and a reference to the places in which the information on each listing can be found.

**index of cooperation:** In fax, the individual scan line length; it should be the same for both the sending and receiving terminals for good quality reproduction.

**information:** Any facts or data which can be used, transferred, or communicated.

**Information management system:** See "IMS."

**information processing:** The receipt, manipulation, storage, or transmission of information.

**information retrieval system:** Short for information storage and retrieval system; a method of identifying and storing data and recovering it when it is needed.

**information system:** The organized acquisition, classification, storage, manipulation, maintenance, and dissemination of information.

**Initial Graphics Exchange Specification:** See "IGES."

**input:** Any data that is entered into a computer to be processed, or the act of entering that data.

**input device:** Any mechanism that converts data into electronic signals that can be used by a computer; examples are a keyboard, mouse, or scanner.

**input media:** The various storage devices used to pass information to a photocomposer for typesetting.

**integrated information system:** An information system that uses more than one technology, such as microcomputers and CD-ROM.

**interface:** A parallel or serial connection between devices, devices and people, people, or systems.

**internal document:** a collection of digitally stored data (text, graphics images, or both) that can be processed by a computer. Compare "external document."

**International Organization for Standardization:** See "ISO."

**International Telecommunications Union:** See "ITU."

**I/O:** Input/output; a transfer of data between the computer's CPU and a peripheral.

**I/O channel:** Input/output channel; a special-purpose computer processor that controls the I/O operations including buffering, formatting, and controlling the flow of data.

**IPI:** Intelligent Peripherals Interface.

**ISDN:** Integrated Services Digital Network; officially defined by CCITT as "a limited set of standard interfaces to a digital communications network." The result is a network that offers end users voice, data, and certain image services on end-to-end digital circuits.

**ISO:** International Organization for Standardization (*Organisation Internationale de Normalisation*).

**ITU:** International Telecommunications Union (*Union Internationale des Telecommunications*) is a 160-member formal treaty organization administered by the United Nations. Compare "CCITT."

**jukebox:** An automatic media handler for optical disks and drives; also called a library. Compare "library."

**K:** Kilo; one thousand.

**Kb:** Kilobits; approximately one thousand bits. Actually, $2^{10}$, or 1,024 bits.

**KB:** Kilobytes; approximately one thousand bytes. Actually, $2^{10}$, or 1,024 bytes."

**Kbps:** Kilobits per second.

**keyword:** A word (or phrase) from the text or title of a document that is descriptive of the contents of that document.

**kilobits:** See "Kb."

**kilobytes:** See "KB."

**LAN:** Local Area Network; networks, generally microcomputer based,that enable users in the same location to use the same programs and equipment such as printers. Compare "WAN."

**laser:** Light Amplification from Stimulated Emission of Radiation; a device for producing light by emission of energy stored in a molecular or atomic system when stimulated by an input signal. A laser beam can be focused with great precision.

**laser printer:** A high-quality, fast (usually eight pages per minute), nonimpact printer that works on the principal of fusing toner to uncharged areas of a paper that has been statically charged. The uncharged areas are created with a laser beam.

**library:** A collection of data files, programs, magnetic tapes, books, or other forms of information. Compare "jukebox."

**line art:** image composed only of black lines on a white background with no shades of gray.

**lpi:** lines per inch.

**M:** Mega; one million.

**m:** Milli; one thousandth or $10^{-3}$.

**magnetic disk:** A rotatable, circular, flat plate that can store data on both sides as a series of positive and negative charges. A hard disk has a metal or glass base and rotates at high speed; a floppy disk has a plastic base. Data are stored on magnetic disks in concentric, circular tracks divided into sectors, and provide fast, random access to any item recorded.

**magnetic media:** Magnetically coated materials used for computer storage. These include cartridges, cassettes, disks, diskettes, magnetic cards, and bubble memories.

**magnetic storage:** Any device that makes use of the magnetic properties of materials for the storage of data.

**magnetic tape:** Ribbon of paper, metal, or plastic, coated or impregnated with magnetic material on which data may be stored in the form of magnetically polarized areas.

**magnetic tape transport:** See "tape drive."

**mainframe:** A large, expensive, powerful computer intended for centralized application.

**mass storage system:** See "MSS."

**Mb:** Megabit; approximately one million bits. Actually, $2^{20}$, or 1,048,576 bits.

**MB:** Megabyte; approximately one million bytes. Actually, $2^{20}$, or 1,048,576 bytes.

**Mbps:** Megabits per second.

**MCAV:** Modified Constant Angular Velocity; a media format with greater data density than Constant Angular Velocity (CAV), but without the performance compromise of Constant Linear Velocity (CLV). Compare "CAV" and "ZCAV."

MCLV: Modified Constant Linear Velocity; Compare "CLV" and "QLV."

media: Properly, plural of "medium," now often used as either singular or plural. See "medium."

media conversion: The process of transferring data from one medium to another, such as from magnetic tape to optical disk.

media-stored format: The symbols that tell a computer how to format text, stored with that text. For example, an underlined title would be preceded with "underline on" and concluded with "underline off."

medium: A substance or object on which data is stored; usually refers to the sensitive coating on a writable device or to the device itself (for example, disk, tape, or card).

megabit: See "Mb."

megabyte: See "MB."

memory: See "storage."

menu: The list of available software functions for selection by the operator, displayed on the computer screen once a software program has been entered.

MHS: Message Handling System; method used to pass electronic mail (E-mail) messages from sender on one network node to a recipient on another node. MHS can handle E-mail messages in the background so that recipient is notified about the arrival of the message but the recipient's current work session is not disrupted.

MHz: Megahertz; a million cycles per second.

migration: The movement of data up and down within a hierarchy of storage devices.

millisecond: See "ms."

MIL-SPEC: Military specification.

MIPS: Million instructions per second; a measure of the speed of a computer.

MMRC: Modified Modified Read Code; raster compression scheme for total two-dimensional coding used in the CCITT Group 4 specification.

mnemonic code: Instructions for the computer written in a form that is easy for the programmer to remember. A program written in mnemonics must be converted to machine code prior to execution.

M-O: Magneto-optic; information stored by local magnetization of a magnetic medium. Reading is performed optically, through rotation of the plane of polarization of probing light via the Faraday effect or Kerr effect.

modem: MOdulator-DEModulator; a device used to convert serial digital data from a transmitting terminal to a signal suitable for transmission over a telephone channel, or to reconvert the transmitted signal to serial digital data for acceptance by a receiving terminal.

modified Huffman code: A method of data compression adopted by CCITT to remove (only) horizontal redundancy from an image. Compare "Huffman code" and "run length encoding."

**Modified Modified Read Code:** See "MMRC."

**Modified Read Code:** See "MRC."

**modulation:** The variation in the value of some parameter characterizing a periodic oscillation. Specifically, variation of some characteristic of a radio wave, called the carrier wave, in accordance with instantaneous values of another wave, called the modulating wave.

**MRC:** Modified Read Code; raster compression scheme for one- and two-dimensional coding used in the CCITT Group 3 specification.

**ms:** Millisecond; one thousandth (0.001) of a second.

**MSP:** An image file format for Microsoft Windows Paint program.

**MSS:** Mass storage system; a system that provides a very large memory capacity, i.e., greater than one terabit. Generally mass storage is relatively inexpensive and is used for backup purposes.

**MVS:** Multiple Virtual Systems; operating environment for IBM mainframes. Compare "VMS."

**n:** Nano; one billionth, or $10^{-9}$.

**National Information Standards Organization:** See "NISO."

**National Institute for Standards and Technology:** See "NIST."

**NCP:** Network Control Program; manages the acquisition, routing, and switching of communication.

**network:** An interconnection of two or more computers for the mutual or individual processing of data to and from a multitude of terminals or stations by using appropriate switching techniques, transmission systems, or miniprocessors.

**network architecture:** The design of system and logic organization and information flow relationships, protocol, and policies for a network.

**NISO:** National Information Standards Organization; helps develop information standards used by libraries.

**NIST:** National Institute of Standard and Technology; the Federal agency charged with developing Federal Information Processing Standards (FIPS) for all other Federal agencies. Compare "FIPS" and "FIPS PUBS."

**node:** In a network, a point of interconnection. Normally, a point at which a number of terminals or tail circuits attach to the network.

**noise:** Unwanted signals present on a medium, either before or after recording.

**ns:** Nanosecond; one billionth of a second. Also abbreviated as "nsec."

**nsec:** See "ns."

**OA:** Office automation; the use of electronic and computer devices in office applications.

**OCR:** Optical character recognition; a process of recognizing symbols, letters of the alphabet, and numbers in printed or bit-mapped form through the use of optical scanning technology. Compare "ICR."

**OD:** Optical disk; a disk that is read from and/or written to by light, generally laser light. Such a disk can store video, audio, or digital data.

**ODD:** Optical data disk. See "OD."

**OEM:** Original equipment manufacturer.

**off-the-shelf:** Generally, professionally written, generic programs that can be purchased ready to use, as opposed to those that must be custom-designed and written for a particular purpose.

**OLTP:** Online transaction processing.

**online computer systems:** Systems where the input data enters the computer directly from the point of origin and/or in which output data is transmitted directly to where it is used.

**operating system:** See "OS."

**optical disks:** See "OD."

**optical file server:** A PC with a high-capacity optical disk acting as a file server. See "file server."

**optical scanning device:** A light source and phototube combined as a single unit for scanning strips of paper or other materials in photoelectric side-register control systems.

**OROM:** Optical Read-Only Memory. See "DATAROM."

**OS:** Operating system; computer programs for expediting, controlling and/or recording computer use by other programs. Used for computer executive systems.

**OSF:** Open Systems Foundation.

**OSI:** Open Systems Interconnection; a mass storage interface standard developed by the International Organization for Standardization.

**output:** Any product of an activity; in computers, the output may be in the form of a display, a report, or a machine-readable product such as a tape or disk.

**output device:** Any piece of equipment that converts computer signals into a form that is readable by humans or by another machine. Examples include a VDT and a printer.

**output media:** Any product produced by an output device, Examples include punched tape or cards, documents, and reports.

**packet-switching network:** A group of geographically scattered devices that exchange packets of data via telecommunication. At the point of transmission a message is divided into fixed-length packets that are sent, automatically, by any route open to the addressed destination. Different packets may be sent by different routes. The message is automatically reassembled at the point of receipt.

*Page Description Language:* See "PDL."

**page memory:** A facsimile machine's capability of retaining a page of data until retrieved through the use of a password.

**page scrolling:** A system's ability to display an entire page and to move forwards or backwards through an entire document.

**pages per minute:** See "ppm."

**parallel interface:** A multiple line channel that supports the simultaneous transmission of groups of bits or bytes. See also "serial transmission" and "parallel printer."

**parallel printer:** In data communications, a printer that accepts each input word as groups of bits sent simultaneously.

**parallel processing:** The simultaneous use of multiple processors to perform a computer operation.

**PC:** Personal Computer; a powerful, desktop computing device that is built with integrated circuit (IC) chips and is designed for an individual user.

**PCL:** Printer Command Language; the proprietary Page Description Language (PDL) used by the Hewlett-Packard LaserJet series of laser printers. PCL uses bit-mapped representation for text fonts. Compare "PostScript."

**PC-to-fax interface:** Refers to an expansion card, or board, that is added to a PC to connect to a communications line and comes with software that allows input at a PC to be converted into format, usually CCITT Group 3, that can be transmitted directly to and output by a fax machine. A more sophisticated PC-to-fax interface allows the PC to receive a fax for output to the PC's monitor or printer.

**PCX:** A bit-mapped image file format created by ZSoft Inc. for its PC Paintbrush program.

**PDL:** Page Description Language; a computer language that allows the user to change a printer's prespecified font, print sizes, and margins, and to create graphics, as well as construct a conventional program.

**pel:** See "pixel."

**peripheral equipment:** Sometimes shortened to peripheral; devices that work in conjunction with a computer but are not part of the computer itself. Examples of peripherals are a disk drive, printer, monitor, card or paper-tape readers or punches, and magnetic tape handlers.

**pixel:** Abbreviation of "picture element," the smallest resolvable dot in an image display. Also abbreviated as "pel."

**point-to-point:** A communication link between two, and only two, pieces of equipment.

**port:** A CPU interface capable of attaching to a modem for communicating with a remote terminal.

**POSIX:** Portable Operating System Interface for Computer Environments; based on UNIX, POSIX is standard for operating systems that can run on a variety of computer platforms.

**PostScript:** A proprietary Page Description Language (PDL) of Adobe Systems; CRT display and printer driver technology that permits quality publishing of text and graphics using desktop computers. Postscript uses scalable outlines of text fonts.

**ppm:** pages per minute; the duty cycle rate used to describe speed of nonimpact printers, usually laser printers, that compose each page separately before ejecting the printed page.

**premastering:** The stage between the original input of data into a system and the creation of the final master file of that data. It may include converting data to a different format, adding error correction codes, or any pre-processing needed to produce such products as CD-ROM, a tape, or a video disk.

**print queue:** an area in computer memory or storage where data to be printed is sent and then processed by the computer in the order received. Depending on the capabilities of the software used to control the print queue, an individual job or file can be removed from the queue without disrupting the other jobs there, or a job with a higher priority can be placed at the front of the queue so it will be printed next.

**printer:** In computers, a peripheral device that produces computer output as printed images on paper.

**protocol:** A formal set of conventions governing the formatting and relative timing of message exchange between two communicating systems.

**PSDN:** Packet-Switched Data Network.

**QLV:** Quantized Linear Velocity; compare "CLV" and "MCLV."

**query:** Computer input from the user requiring a computer response. A user may count items in a database, compile them, or list them in different ways through the use of queries.

**query language:** A command language used to search and retrieve information. Usually part of a database management system.

**queue:** See "print queue."

**RAM:** Random Access Memory; general-purpose reconfigurable computer chips that hold data only as long as the computer is on. If the data in RAM are not stored to a durable medium, such as magnetic or optical disks, they are lost when the computer is turned off. Compare "ROM."

**raster:** The scanning lines that collectively form the image or graphics output on a VDT. Compare "vector."

**raster image:** A technique in which an image is captured or displayed. For instance, in order to display an image on a VDT, the system must reconvert a series of digital pixel values into an analog signal for display. Compare "vector image."

**real time:** Pertains to the actual time during which a physical process transpires; generally implies an equivalency to human perception of elapsed time.

**records processing:** The rearrangement of records or parts of records in a file to produce a report for a particular purpose.

**reference database:** A database that does not contain full text, but that does provide the user with information on where to find additional details or primary sources.

**referential integrity:** A process that ensures that each data item in a database belongs to a defined group.

**relational database:** A database constructed with logical links between fields in the various table-like files, but not between files; designed to reduce data redundancy and to make it possible to collect data by related fields, from multiple files.

**remote access:** The use of a computer from a geographically distant terminal.

**report:** Computer-generated output printed for the user's information.

**RES:** An image file format developed by Xerox Systems Inc.

**resolution:** See "image resolution."

**RFT:** Revisable Form Text.

**RIFF:** Raster Image File Format; a format used for storing scanned grayscale images. Compare "TIFF."

**RISC:** Reduced Instruction Set Computer.

**ROM:** Read-Only Memory; compare "RAM."

**RS-232-C:** Recommended Standard developed by the Electronic Industries Association for interfaces between dedicated word processors, computers, modems, and transmission lines; basically a plug of 25 pins, 17 of which are defined. Unbalanced; supports a transmission of 20,000 bps.

**RS-422-A:** Recommended Standard developed by the Electronic Industries Association for single-pathway computer communications. It offers greater speed and less noise than RS-232-C.

**RS-449:** Recommended Standard developed by the Electronic Industries Association for interfaces between word processors, computers, modems, and transmission lines The 37-pin plug supports a transmission speed of 2 Mbps. It comes in two versions: RS-422, which is balanced, and RS-423, which is unbalanced.

**RS-488:** Recommended Standard developed by Hewlett-Packard for a test and measurement device sometimes used as an interface in data communications. This RS has been modified and approved by the Institute of Electrical and Electronics Engineers (IEEE).

**run length encoding:** The basis for most of the data compression methods used in digital representation of images; based on transmitting numbers describing the lengths of white and black regions of an image rather than sending separately each black or white pixel. Compare "modified Huffman code."

**SAA:** Systems Application Architecture; a standard promoted by IBM that allows applications and their data to be used on a range of computing platforms, including minicomputers and mainframes. See "CUA."

**scaling:** A transformation algorithm converting one bit-map density into a bit-map of another density. Scaling usually involves enlarging or contracting an image or font outline.

**scanner:** A device that resolves a two-dimensional object, such as a business document, into a stream of bits by raster scanning and quantization.

**scanning rate:** The number of lines per minute that a facsimile machine sends or receives.

**SCSI:** Small Computer System Interface; an intelligent device allowing devices to be connected serially. Commonly pronounced "scuzzy."

**SDLC:** Synchronous Data Link Control; IBM standard protocol that supersedes bisynchronous (BSC) transmission.

**serial interface:** The computer connection through which words are transmitted to a printer bit by bit as opposed to a word at a time. Compare "parallel interface."

**serial transmission:** The most common transmission mode; in serial, information bits are sent sequentially on a single data channel.

**SGML:** Standard Generalized Markup Language; a standard method of tagging an electronic document file for indicating typesetting and layout formats. Compare "CALS."

**sheet feeder:** A printer peripheral controlled by a microcomputer. The feeder can select paper from either of two trays, or select envelopes, position them for printing, and receive the ejected paper.

**SNA:** System Network Architecture; the unified software method promoted by IBM for communications access and use across computing environments and platforms.

**software:** A general term for computer programs and related documentation. Compare "firmware," "system software," and "application software."

**source document:** An document from which data is entered into a computer. Compare "external document."

**SQL:** Structured Query Language; language for manipulating relational database information.

**storage:** The part of a computer that holds data and returns data on demand; also called memory.

**storage capacity:** The number of bytes of data that can be held in main memory or in other devices connected to the computer system; generally expressed in kilobytes.

**Structured Query Language:** See "SLQ."

**system:** Any combination of things coordinated to accomplish some function.

**system software:** Programs that perform control functions such as database management, or the control of the operating system or a network.

**T:** Tera; one trillion; $10^{12}$, or 1,000,000,000,000.

**tail circuit:** The circuit at the end of a chain of circuits in a network.

**tape drive:** The mechanism that controls the movement of a magnetic tape past heads that read, erase, or write. Control includes the tape's tension, torque, and speed.

**Technical and Management Information System:** See "TMIS."

**telecommunication:** Communication between two or more distant points via signals, commonly electric, electromagnetic radiation, or light.

**terabit:** Approximately one trillion bits; actually, $2^{40}$, or 1,009,511,627,776 bits.

**teleprocessing:** Manipulation by a computer of data received via telecommunication from a distant source or for a distant user.

**text files:** File of primary source natural language documents stored in code, often ASCII.

**throughput:** The volume of work or information flowing through a system, particularly meaningful in information storage and retrieval systems, in which throughput is measured in units such as accesses per hour.

**TIFF:** Tagged Image File Format; an industry standard, originally developed by Aldus Corporation, that is used tags (labels) to define the exact structure of graphics images capture by a digital scanner to be processed.

**timesharing:** A method of computer operation that allows several interactive terminals to use one computer. Although the terminals are actually served in sequence, the high speed of the computer makes it appear as if all terminals were being served simultaneously.

**TIMS:** Text Information Management System.

**TMIS:** Technical and Management Information System; an information management system developed by Boeing for NASA's Space Station program.

**Token Ring:** A Local Area Network developed by IBM in which a terminal may transmit or receive a message only when it has a token that is passed, ring-like, from terminal to terminal. The token is a unique group of bits used to frame or identify a message. Only one message may be sent at a time.

**transfer rate:** The rate at which data is transferred to or from a device; especially the reading or writing rate of a storage peripheral. Usually expressed in kilobits or megabits per second (Kbps or Mbps).

**turnaround time:** The actual time required to reverse the direction of transmission from sender to receiver or vice versa when using a half-duplex circuit. Time is required for line propagation effects, modem timing, and computer reaction.

**turnkey system:** A complete computer system designed for a user, with all necessary hardware installed, programs and documentation complete and in place, ready for the customer to "turn the key."

**type font:** A set of characters that are all one point size and type style.

**type size:** Type size of a specific font, is measured in points, 72 to the inch or 0.3515 millimeters each, from the top of a capital letter to the bottom of the descender. A descender is the part of a letter that descends below the line. The width is in proportion to the height. Two standard sizes of type used for typewriters are pica and elite. In printing, pica letters are 12 points high; a typewriter prints 10 pica characters to the inch. Elite is smaller; a typewriter prints 12 elite characters to the inch.

**unattended operation:** Computerized transmission of data or printing performed without an operator in attendance.

**UNIX:** A multitasking, multiuser operating system for computers and workstations developed by AT&T and subsequently placed into the public domain. Each computer system vendor has created its own variation of UNIX, such as Apple's A/UX for Macintosh and IBM's AIX for the RISC System/6000 (RS/6000), with the result that no single variation of UNIX is the standard and an application created under UNIX on one platform does not necessarily operate without modification on other platform.

**upload:** To send data or a program from one computer to another (usually larger) one, as from a minicomputer to a mainframe.

**VDT:** Video display terminal; the monitor screen that shows the text and graphics images processed by a computer or workstation. Compare "CRT."

**vector:** A line from one given point to another in a three-dimensional space. Compare "raster."

**vector image:** An image, usually a drawing, using specific *X-Y* coordinates with a precise starting point, length, and direction. Compare "raster image."

**video:** Pertains to the images produced by a television or computer. While television/video has a single standard in the United States, computer/videos vary in their bandwidth and spectrum requirements.

**VMS:** Virtual Memory Operating System; operating system for Digital Equipment Corporation (DEC) minicomputers. Compare "MVS."

**VTAM:** Virtual Telecommunications Access Method.

**WAN:** Wide Area Network; generally, a WAN requires remote data communication facilities between all or some of its points, whereas a Local Area Network (LAN) is entirely local and does not. Multiple remote LANs can be connected into a WAN. Compare "LAN."

**wideband:** A generic term defining a type of telecommunications link using a bandwidth beyond voice grade (below 9,600 bps) providing data transmission speeds of up to several million bits per second.

**Winchester disk:** A nonremovable hard, magnetic storage disk sealed in a unit, with an access arm, read/write heads, and often with other disks. Named Winchester for its original dual 30MB configuration (30-30 being a Winchester rifle specification). See "hard disk."

**workstation:** (1) synonym for a minicomputer; (2) the input/output equipment at which an operator works; or, a PC, microcomputer, or minicomputer and any hardware, such as a keyboard, display unit, printer, media drives, or modem, necessary to do a particular job; (3) any external factor such as the desk, chair, lighting, noise, or layout that can affect a person's ability to do assigned work in a particular location.

**WORM:** Write Once, Read Many times; a type of nonerasable optical disk storage—that is, once written to a WORM disk, data cannot be erased. Compare "erasable."

**write-once:** *See "WORM."*

**WYSIWYG:** What you see is what you get.

**X.25:** CCITT standard governing interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for terminals operating in the packet-switched mode on public data networks.

**X Window System:** A Graphical User Interface (GUI) developed by the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts, that is rapidly gaining acceptance as a windowing environment standard for UNIX-based computers and workstations.

**ZCAV:** Zoned Constant Angular Velocity; Compare "CAV" and "MCAV."

## E.2.    GLOSSARY REFERENCES

"CD-ROM Publishing: Glossary of Frequently Used Technical Terms." Knowledge Access International. *Laserdisk Professional*, January 1990, pp. 51–54.

*Chambers Science and Technology Dictionary, General Edition*. Peter M. B. Walker. New York: Chambers. 1988.

*Computer Dictionary and Handbook*. Charles J. Sippl. Howard W. Sams & Company Inc., publisher. New York: The Bobbs-Merrill Company Inc. 1967.

*Computer Glossary, The: The Complete Illustrated Desk Reference, Fourth Edition*. Alan Freedman. American Management Association, publisher. Point Pleasant, PA: The Computer Language Company Inc. 1989.

*Dictionary of Computer Terms, Second Edition*. Michael Covington, Ph.D., and Douglas Downing, Ph.D. New York: Barrons. 1989.

"Glossary." *Black Box Catalog*, May 1990, pp. 82–97. The Black Box Corporation, Pittsburgh, PA.

*IEEE Standard Dictionary of Electrical and Electronics Terms, Fourth Edition*. Frank Jay, Editor-in-Chief. The Institute of Electrical and Electronics Engineers Inc., publisher. New York: John Wiley and Sons Inc. 1988.

*Inside X.25: A Manager's Guide*. Sherman K. Schlar. New York: McGraw-Hill Publishing Company. June 1990.

*McGraw-Hill Dictionary of Scientific and Technical Terms, Fourth Edition*. Sybil P. Parker, Edition-in-Chief. New York: McGraw-Hill Publishing Company. 1989.

"User's Guide and Glossary." *Datapro Reports on Document Imaging Systems*, February 1990, pp. 7–33. Datapro Research, Delran, NJ.

*Van Nostrand Reinhold Dictionary of Information Technology, Third Edition*. Dennis Langley and Michael Shain. New York: Van Nostrand Reinold. 1989