

***Exhibit A-1***



THE SEDONA  
CONFERENCE® GLOSSARY:  
*E-Discovery &  
Digital Information  
Management*

A Project of The Sedona Conference®  
Working Group on Electronic Document  
Retention & Production (WG1) RFP+ Group

MAY 2005 VERSION



# THE SEDONA CONFERENCE® GLOSSARY FOR E-DISCOVERY AND DIGITAL INFORMATION MANAGEMENT

## *Co-Editors:*

Conor R. Crowley, Esq.  
Labaton, Sucharow & Rudoff, LLP  
New York, NY

Sherry B. Harris  
Hunton & Williams  
Richmond, VA

## *Contributing Editors:*

Matthew I. Cohen, Esq.  
Skadden, Arps, Slate, Meagher & Flom  
New York, NY

Anne Kershaw, Esq.  
Kershaw Attorneys & Consultants, PC  
Tarrytown, NY

Mark V. Reichenbach  
Milberg Weiss Bershad & Schulman  
New York, NY

RFP+ Vendor Panel (see  
[www.thesedonaconference.org](http://www.thesedonaconference.org) for a listing of  
the RFP+ Vendor Panel

Copyright © 2005 The Sedona Conference®  
All Rights Reserved

## REPRINT REQUESTS

Requests for reprints or reprint information should be directed to  
Richard Braman, Executive Director of The Sedona Conference,  
at [tsc@sedona.net](mailto:tsc@sedona.net) or 1-866-860-6600.

The opinions expressed in this publication, unless otherwise attributed, represent consensus views of the Editors and the members of The Sedona Conference® Working Group on Electronic Document Retention and Protection, RFP+ Group. They do not necessarily represent the views of any of the individual participants or editors or their employers, clients, or any other organizations to which any of the participants belong nor do they necessarily represent official positions of The Sedona Conference®.

---

The logo consists of the letters 'WGS' in a bold, sans-serif font. The 'W' and 'G' are connected at the top, and the 'S' is slightly larger and positioned to the right. A small 'SM' trademark symbol is located at the top right of the 'S'.

Copyright © 2005  
The Sedona Conference®

Visit [www.thesedonaconference.org](http://www.thesedonaconference.org)

---

## *Note to Users*

---

The Sedona Conference® Glossary is published as a tool to assist in the understanding and discussion of electronic discovery and electronic information management issues; it is not intended to be an all-encompassing replacement of existing technical glossaries published by ARMA International ([www.arma.org](http://www.arma.org)), American National Standards Institute ([www.ansi.org](http://www.ansi.org)), International Organization for Standardization ([www.iso.org](http://www.iso.org)), U.S. National Archives & Records Administration ([www.archives.gov](http://www.archives.gov)) and other professional organizations. As with all of our publications, your comments are welcome. Please forward them to us at [tsc@sedona.net](mailto:tsc@sedona.net).

*Richard G. Braman*  
Executive Director  
The Second Conference®  
Sedona, AZ  
USA  
May, 2005

## Commonly Used Terms for E-Discovery and Digital Information Management<sup>1</sup>

30(b)(6) : Under Federal Rule of Civil Procedure 30(b)(6), a corporation, partnership, association, or governmental agency is subject to the deposition process, and to provide one or more witnesses to “testify as to matters known or reasonably available to the organization” without compromising attorney-client privilege communications or work product. It is not unusual for the 30(b)(6) topics to be directed toward the discovery process, including procedures for preservation, collection, chain of custody, processing, review, and production. Early in the litigation, when developing a discovery plan, particularly with regard to electronic discovery, a party should be mindful of the obligation to provide one or more 30(b)(6) witnesses should the request be made by another party to the litigation, and include this contingency in the discovery plan.

*Ablate*: Describes the process by which laser-readable “pits” are burned into the recorded layer of optical discs, DVD-ROMs and CD-ROMs.

*Ablative*: Unalterable data. *See Ablate*.

*Acetate-base film*: A safety film (ANSI Standard) substrate used to produce microfilm.

*ACL (Access Control List)*: A security type used by Lotus Notes developers to grant varying levels of access and user privileges within Lotus Notes databases.

**Active Data**: Active Data is information residing on the direct access storage media (disc drives or servers) of computer systems, which is readily visible to the operating system and/or application software with which it was created and immediately accessible to users without restoration or reconstruction.

**Active Records**: Active Records are those Records related to current, ongoing or in-process activities and are referred to on a regular basis to respond to day-to-day operational requirements. An active record resides in native application format and is accessible for purposes of business processing with no restrictions on alteration beyond normal business rules. *See Inactive Records*.

*ADC*: Analog to Digital converter. Converts analog data to a digital format.

*Address*: Addresses using a number of different protocols are commonly used on the Internet. These addresses include email addresses (Simple Mail Transfer Protocol or SMTP), IP (Internet Protocol) addresses and URLs (Uniform Resource Locators), commonly known as Web addresses.

*ADF*: Automatic Document Feeder. This is the means by which a scanner feeds the paper document.

*AIIM*: The Association for Information and Image Management – focused on electronic imaging.

---

<sup>1</sup> Terms previously defined in *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production* or *The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age* are in **bold**.

**Algorithm:** A detailed formula or set of steps for solving a particular problem. To be an algorithm, a set of rules must be unambiguous and have a clear stopping point.

**Aliasing:** When computer graphics output has jagged edges or a stair-stepped, rather than a smooth, appearance when magnified. The graphics output can be smoothed using anti-aliasing algorithms.

**Alphanumeric:** Characters composed of letters, numbers (and sometimes punctuation marks). Excludes control characters.

**Ambient Data:** See Residual Data.

**Analog:** Data in an analog format is represented by continuously variable, measurable, physical quantities such as voltage, amplitude or frequency. Analog is the opposite of digital.

**Annotations:** The changes, additions, or editorial comments made or applicable to a document - usually an electronic image file - using electronic sticky notes, highlighter, or other electronic tools. Annotations should be overlaid and not change the original document.

**ANSI:** American National Standards Institute - a private, non-profit organization that administers and coordinates the U.S. voluntary standardization and conformity assessment system.

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

**Application:** An application is a collection of one or more related software programs that enable an end-user to enter, store, view, modify, or extract information from files or databases. The term is commonly used in place of “program,” or “software.” Applications may include word processors, Internet browsing tools, spreadsheets, email clients, personal information managers (contact information and calendars), and other databases.

**Architecture:** The term architecture refers to the hardware, software or combination of hardware and software comprising a computer system or network. The term “open architecture” is used to describe computer and network components that are more readily interconnected and interoperable. Conversely, the term “closed architecture” describes components that are less readily interconnected and interoperable.

**Archival Data:** Archival Data is information an organization maintains for long-term storage and record keeping purposes, but which is not immediately accessible to the user of a computer system. Archival data may be written to removable media such as a CD, magneto-optical media, tape or other electronic storage device, or may be maintained on system hard drives. Some systems allow users to retrieve archival data directly while other systems require the intervention of an IT professional.

**Archive, Electronic Archive:** Archives are long-term repositories for the storage of records. Electronic archives preserve the content, prevent or track alterations, and control access to electronic records.

**ARMA International:** a not-for-profit association and recognized authority on managing records and information – paper and electronic - [www.arma.org](http://www.arma.org).

**Artificial Intelligence (AI):** The subfield of computer science concerned with the concepts and methods of symbolic inference by computer and symbolic knowledge representation for use in making inferences - an attempt to model aspects of human thought on computers. It is also sometimes defined as trying to

solve by computer any problem once believed to be solvable only by humans. AI is the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience. It attempts to approximate the results of human reasoning by organizing and manipulating factual and heuristic knowledge. Areas of AI activity include expert systems, natural language understanding, speech recognition, vision, and robotics.

*ASCII (American Standard Code for Information Interchange):* Pronounced “ask-ee,” ASCII is a non-proprietary text format built on a set of 128 (or 255 for *extended ASCII*) alphanumeric and control characters. Documents in ASCII format consist of only text with no formatting and can be read by most computer systems.

*Aspect Ratio:* The relationship of the height and width of any image. The aspect ratio of an image must be maintained to prevent distortion.

**Attachment:** An attachment is a record or file associated with another record for the purpose of retention or transfer. There may be multiple attachments associated with a single “parent” or “master” record. In many records and information management programs the attachments and associated record are managed and processed as a single unit. In common use, this term refers to a file (or files) associated with an email for retention and storage as a single message unit.

**Attribute:** An attribute is a characteristic of data that sets it apart from other data, such as location, length, or type. The term attribute is sometimes used synonymously with “data element” or “property”.

**Audit Trail** – In computer security systems, a chronological record of when users logged in, how long they were engaged in various activities, what they were doing, and whether any actual or attempted security violations occurred. An audit trail is an automated or manual set of chronological records of system activities that may enable the reconstruction and examination of a sequence of events and/or changes in an event.

**Author or Originator:** The author of a document is the person, office or designated position responsible for its creation or issuance. In the case of a document in the form of a letter, the author or originator is usually indicated on the letterhead or by signature. In some cases, the software application producing the document may capture the author’s identity and associate it with the document. For records management purposes, the author or originator may be designated as a person, official title, office symbol, or code.

*AVI (Audio-Video Interleave):* A Microsoft standard for Windows animation files that interleaves audio and video to provide medium quality multimedia.

*Backbone:* The top level of a hierarchical network. It is the main channel along which data is transferred.

*Backfiles:* Existing paper or microfilm files.

*Backup:* To create a copy of data as a precaution against the loss or damage of the original data. Many users backup their files, and most computer networks utilize automatic backup software to make regular copies of some or all of the data on the network.

**Backup Data:** An exact copy of system data which serves as a source for recovery in the event of a system problem or disaster. Backup Data is generally stored separately from Active Data on portable media. Backup Data is distinct from Archival Data in that Backup Data may be a copy of Active Data,

but the more meaningful difference is the method and structure of storage which impact its suitability for certain purposes.

**Backup Tape Recycling:** Backup Tape Recycling describes the process whereby an organization's backup tapes are overwritten with new data, usually on a fixed schedule determined jointly by records management, legal, and IT sources. For example, the use of nightly backup tapes for each day of the week with the daily backup tape for a particular day being overwritten on the same day the following week; weekly and monthly backups being stored offsite for a specific period of time before being placed back in the rotation.

**Backup Tapes:** Magnetic tapes used to store copies of data, for use when restoration or recovery of data is required. Data on backup tapes are generally recorded and stored sequentially, rather than randomly, meaning in order to locate and access a specific file or data set, all data on the tape preceding the target must first be read, a time-consuming and inefficient process. Backup tapes typically use data compression, which increases restoration time and expense, given the lack of uniform standards governing data compression.

**Bandwidth:** The amount of information or data that can be sent over a network connection in a given period of time. Bandwidth is usually stated in kilobits per second (kbps) or megabits per second (mps).

**Bar Code:** A small pattern of vertical lines that can be read by a laser or an optical scanner. In records management and electronic discovery, bar codes are often affixed to specific records for indexing, tracking and retrieval purposes.

**Batch Processing:** The processing of a large amount of data, or multiple records, in a single step.

**Bates Number:** Sequential numbering used to track documents and images in production data sets, where each page is identified by a unique production number. Often used in conjunction with a suffix or prefix to identify the producing party, the litigation, or other relevant information. *See also* Production Number.

**Baud Rate:** The number of times per second a communications channel changes the carrier signal it sends on a phone line. A 2400-baud modem changes the signal 2400 times a second.

**BBS (Bulletin Board System):** A computer system or service that users access to participate in electronic discussion groups, post messages and/or download files.

**BCS:** Boston Computer Society, one of the first associations of PC/Apple users and one of the largest and most active.

**Beginning Document Number or BegDoc#:** The Bates Number identifying the first page of a document or record.

**Bibliographical/Objective Coding:** Extracting objective information from electronic documents such as date created, author/recipient/copies, and associating the information with a specific electronic document.

**Binary:** The Base 2 numbering system used in digital computing which represents all numbers using combinations of zero and one.



*BIOS (Basic Input Output System):* The set of user-independent computer instructions stored in a computer's ROM, immediately available to the computer when the computer is turned on. BIOS information provides the code necessary to control the keyboard, display screen, disc drives and communication ports in addition to handling certain miscellaneous functions.

*Bit Map:* A Bit Map provides information on the placement and color of individual bits and allows the creation of characters or images by creating a picture composed of individual bits (pixels).

*Bit Stream Back-up:* A Bit Stream Back-up is a sector-by-sector/bit-by-bit copy of a hard drive. A Bit Stream Back-up is an exact copy of a hard drive, preserving all latent data in addition to the files and directory structures. Bit Stream Back-up may be created using applications such as Encase, SnapBack and Ghost. See Forensic Copy.

*Bit:* A bit (binary digit) is the smallest unit of computer data. A bit consists of either 0 or 1. There are eight bits in a byte.

*Bi-Tonal:* A bi-tonal image uses only black and white.

*BMP:* A Windows file format for storing bit map images.

*Bookmark:* A link to a Web site or page previously visited.

*Boolean Search:* Boolean Searches use the logical operators "and", "or" and "not" to include or exclude terms from a search.

*Boot Sector:* The very first sector on a hard drive which contains the computer code (boot strap loader) necessary for the computer to start up and the partition table describing the organization of the hard drive.

*Boot:* To start up or reset a computer.

*BPI (Bits Per Inch):* BPI measures data densities in disc and magnetic tape systems.

*Bps:* bits per second.

*Broadband:* Communications of high capacity and usually of multimedia content.

*Browser:* An application, such as Internet Explorer or Netscape Navigator, used to view and navigate the World Wide Web and other Internet resources.

*Bug:* A problem with computer software or hardware that causes it to malfunction or crash.

*Burn:* The process of a creating a copy of information onto a CD or DVD.

*Bus:* A parallel circuit that connects the major components of a computer, allowing the transfer of electric impulses from one connected component to any other.

*Business Process Outsourcing:* Business process outsourcing occurs when an organization turns over the management of a business function, such as accounts payable, purchasing, payroll or information technology to a third party.

*Byte (Binary Term):* A Byte is the basic measurement of most computer data and consists of 8 bits. Computer storage capacity is generally measured in bytes. Although characters are stored in bytes, a few bytes are of little use for storing a large amount of data. Therefore, storage is measured in larger increments of bytes. *See* Kilobyte, Megabyte, Gigabyte, Terabyte, Petabyte and Exabyte.

*Cache:* A dedicated, high speed storage location which can be used for the temporary storage of frequently used data. As data may be retrieved more quickly from cache than the original storage location, cache allows applications to run more quickly. Web site contents often reside in cached storage locations on a hard drive.

*Caching:* The temporary storage of frequently-used data to speed access. *See also* Cache.

*Case De-Duplication:* Eliminates duplicates to retain only one copy of each document per case. For example, if an identical document resides with three custodians, only the first custodian's copy will be saved. *See* De-Duplication.

*Catalog:* *See* Index.

*CCD (Charge Coupled Device):* A computer chip the output of which correlates with the light or color passed by it. Individual CCDs or arrays of these are used in scanners as a high-resolution, digital camera to read documents.

*CCITT Group 4:* A lossless compression technique/format that reduces the size of a file, generally about 5:1 over RLE and 40:1 over bitmap. CCITT Group 4 compression may only be used for bi-tonal images.

*CCITT:* Consultative Committee for International Telephone & Telegraphy. Sets standards for phones, faxes, modems etc. The standard exists primarily for fax documents.

*CDPD (Cellular Digital Packet Data):* A data communication standard utilizing the unused capacity of cellular voice providers to transfer data.

*CD-R (Compact Disc Recordable):* A CD-ROM on which a user may permanently record data once using a CD Burner.

*CD-RW (Compact Disc Re-Writable):* A CD-ROM on which a user may record data multiple times.

*CD-ROM:* *See* Compact Disc.

*Centronics Interface:* A parallel interface standard for connecting printers and other devices to computers.

*Certificate:* Digital signature combining data verification and encryption key. *See* PKI Digital Signature.

*CGA (Color Graphics Adapter):* *See* Video Graphics Adapter (VGA).

*Chaffwinnowing:* Advanced encryption technique involving data dispersal and mixing.

*Chain of Custody:* Documentation and testimony regarding the possession, movement, handling and location of evidence from the time it is obtained to the time it is presented in court; used to prove that

evidence has not been altered or tampered with in any way; necessary both to assure admissibility and probative value.

*Character Treatment:* The use of all caps or another standard form of treating letters in a coding project.

*CIE (Commission International de l'Eclairage):* The international commission on color matching and illumination systems.

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

*Cinepak:* A compression algorithm; *see* MPEG.

*CITIS (Contractor Integrated Technical Information Service):* The Department Of Defense now requires contractors to have an integrated electronic document image and management system.

*Client/Server:* An architecture whereby a computer system consists of one or more server computers and numerous client computers (workstations). The system is functionally distributed across several nodes on a network and is typified by a high degree of parallel processing across distributed nodes. With client-server architecture, CPU intensive processes (such as searching and indexing) are completed on the server, while image viewing and OCR occur on the client. This dramatically reduces network data traffic and insulates the database from workstation interruptions.

*Client:* Any computer system that requests a service of another computer system. A workstation requesting the contents of a file from a file server is a client of the file server. *See* Thin Client.

*Clipboard:* A holding area that temporarily stores information copied or cut from a document.

*Cluster (File):* The smallest unit of storage space that can be allocated to store a file on operating systems that use a file allocation table (FAT) architecture. Windows and DOS organize hard discs based on Clusters (also known as allocation units), which consist of one or more contiguous sectors. Discs using smaller Cluster sizes waste less space and store information more efficiently.

*Cluster (System):* A collection of individual computers that appear as a single logical unit. Also referred to as matrix or grid systems.

*Cluster bitmaps:* Used in NTFS to keep track of the status (free or used) of clusters on the hard drive.

*CMYK:* Cyan, Magenta, Yellow and Black. A subtractive method used in four color printing and Desktop Publishing.

*Coding:* Automated or human process through which documents are examined and evaluated using pre-determined codes, and the results of those comparisons are logged. Coding usually identifies names, dates, and relevant terms or phrases. Coding may be structured (limited to the selection of one of a finite number of choices), or unstructured (a narrative comment about a document). Coding may be objective, i.e., the name of the sender or the date, or subjective, i.e., evaluation as to the relevancy or probative value of documents.

*COLD (Computer Output to Laser Disc):* A computer programming process that outputs electronic records and printed reports to laser disc instead of a printer.

**COM (Computer Output to Microfilm):** A process that outputs electronic records and computer generated reports to microfilm.

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

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

**Comma Separated Value (CSV):** A record layout that separates data fields/values with a comma and typically encloses data in quotation marks.

**Compact Disc (CD):** A type of optical disc storage media, compact discs come in a variety of formats. These formats include CD-ROMs (“CD Read-Only Memory”) that are read-only; CD-Rs (“CD Recordable”) that are write to once and are then read-only; and CD-RWs (“CD Re-Writable”) that can be written to multiple times.

**Compliance Search:** The identification of relevant terms and/or parties in response to a discovery request.

**Component Video:** Separates video into luminosity and color signals that provide the highest possible signal quality.

**Composite Video:** Combines red, green, blue and synchronization signals into one video signal so that only one connector is required; used by most TVs and VCRs.

**Compression Ratio:** The ratio of the size of an uncompressed file to a compressed file, e.g., with a 10:1 compression ratio, a 1 MB file can be compressed to 100 KB.

**Compression:** Compression algorithms such as Zip and RLE reduce the size of files saving both storage space and reducing bandwidth required for access and transmission. Data compression is widely used in backup utilities, spreadsheet applications and database management systems. Compression generally eliminates redundant information and/or predicts where changes will occur. “Lossless” compression techniques such as Zip and RLE preserve the integrity of the input. Coding standards such as JPEG and MPEG employ “lossy” methods which do not preserve all of the original information, and are most commonly used for photographs, audio, and video.

**Computer Forensics:** Computer Forensics (in the context of this document, “forensic analysis”) is the use of specialized techniques for recovery, authentication and analysis of electronic data when an investigation or litigation involves issues relating to reconstruction of computer usage, examination of residual data, authentication of data by technical analysis or explanation of technical features of data and computer usage. Computer forensics requires specialized expertise that goes beyond normal data collection and preservation techniques available to end-users or system support personnel, and generally requires strict adherence to chain-of-custody protocols. *See also* Forensics and Forensic Copy.

**Computer:** Includes but is not limited to network servers, desktops, laptops, notebook computers, mainframes and PDAs (personal digital assistants).

**Concept Search:** Searching electronic documents to determine relevance by analyzing the words and putting search requests in conceptual groupings so the true meaning of the request is considered. Concept searching considers both the word and the context in which it appears to differentiate between concepts such as diamond (baseball) and diamond (jewelry).

*Content Comparison*: A method of de-duplication that compares file content or output (to image or paper) and ignores metadata. *See* De-Duplication.

*Contextual Search*: The process of returning electronic evidence to its true context: when created, by whom, for what purpose, etc.

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

*Convergence*: Integration of computing, communications and broadcasting systems.

*Cookie*: A message given to a Web browser by a Web server. The browser stores the message in a text file. The message is then sent back to the server each time the browser requests a page from the server. The main purpose of cookies is to identify users and possibly prepare customized Web pages for them.

*Corrupted File*: A file damaged in some way, such as by a virus, or by software or hardware failure, so that it cannot be read by a computer.

*COTS (Commercial Off-the-Shelf)*: Hardware or software products that are commercially manufactured, ready-made and available for use by the general public without the need for customization.

*CPI*: Characters Per Inch.

*CPU (Central Processing Unit)*: The primary silicon chip that runs a computer's operating system and application software. It performs a computer's essential mathematical functions and controls essential operations.

*CRC (Cyclical Redundancy Checking)*: Used in data communications to create a checksum character at the end of a data block to ensure integrity of data transmission and receipt.

*CRM (Customer Relationship Management)*: programs that help manage clients and contacts. Used in larger companies. Often a significant repository of sales, customer, and sometimes marketing data.

*Cross-Custodian De-Duplication*: Culls a document to the extent multiple copies of that document reside within different custodians' data sets. *See* De-Duplication.

*CRT (Cathode Ray Tube)*: The picture tube of a computer monitor or television.

*Cryptography*: Technique to scramble data to preserve confidentiality or authenticity.

*Cull (verb)*: To remove a document from the collection to be produced or reviewed. *See* Data Filtering, Harvesting.

*Custodian*: Person having control of a network, computer or specific electronic files.

*Custodian De-Duplication*: Culls a document to the extent multiple copies of that document reside within the same custodian's data set. *See* De-Duplication.

*Customer-Added metadata*: *See* User-Added Metadata.

*CYAN*: Cyan-colored ink reflects blue and green and absorbs red.

*Cylinder*: The set of tracks on both sides of each platter in the hard drive that is located at the same head position.

*DAC (Digital to Analog Converter)*: Converts digital data to analog data.

*DAD (Digital Audio Disc)*: Another term for compact disc.

*DAT (Digital Audio Tape)*: A magnetic tape generally used to record audio but can hold up to 40 gigabytes (or 60 CDs) of data if used for data storage. Has the disadvantage of being a serial access device. Often used for backup.

*Data*: Any information stored on a computer. All software is divided into two general categories: data and programs. Programs are collections of instructions for manipulating data. In database management systems data files are the files that store the database information. Other files, such as index files and data dictionaries, store administrative information, known as metadata.

*Data Collection*: See Harvesting.

*Data Element*: A combination of characters or bytes referring to one separate piece of information, such as name, address, or age.

*Data Extraction*: The process of retrieving data from documents (hard copy or electronic). The process may be manual or electronic.

*Data Field*: See Field.

*Data Filtering*: The process of identifying for extraction specific data based on specified parameters.

*Data Formats*: The organization of information for display, storage or printing. Data is maintained in certain common formats so that it can be used by various programs, which may only work with data in a particular format, e.g. PDF, html.

*Data Harvesting*: See Harvesting.

*Data Mining*: Data mining generally refers to techniques for extracting summaries and reports from an organization's databases and data sets. In the context of electronic discovery, this term often refers to the processes used to cull through a collection of electronic data to extract evidence for production or presentation in an investigation or in litigation.

*Data Set*: A named or defined collection of data. See also Production Data Set and Privilege Data Set.

*Data Verification*: Assessment of data to ensure it has not been modified. The most common method of verification is hash coding by some method such as MD5. See also Digital Fingerprint and File Level Binary Comparison and Hash Coding.

**Database Management System (DBMS)**: A software system used to access and retrieve data stored in a database.

**Database**: In electronic records a database is a set of data elements consisting of at least one file, or of a group of integrated files, usually stored in one location and made available to several users. In computing databases are sometimes classified according to their organizational approach with the most

prevalent approach being the relational database - a tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. Another popular organizational structure is the distributed database which can be dispersed or replicated among different points in a network. Computer databases typically contain aggregations of data records or files, such as sales transactions, product catalogs and inventories, and customer profiles. SQL (Structured Query Language) is a standard computer language for making interactive queries from and updates to a database.

*Daubert (challenge):* *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), addresses the admission of scientific expert testimony to ensure that the testimony is reliable before considered for admission pursuant to Rule 702. The court assesses the testimony by analyzing the methodology and applicability of the expert's approach. Faced with a proffer of expert scientific testimony, the trial judge must determine first, pursuant to Rule 104(a), whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact at issue. This involves preliminary assessment of whether the reasoning or methodology is scientifically valid and whether it can be applied to the facts at issue. *Daubert* suggests an open approach and provides a list of four potential factors: (1) whether the theory can be or has been tested; (2) whether the theory has been subjected to peer review or publication; (3) known or potential rate of error of that particular technique and the existence and maintenance of standards controlling the technique's operation; and (4) consideration of general acceptance within the scientific community. 509 U.S. at 593-94.

*Decryption:* Transformation of encrypted (or scrambled) data back to original form.

**De-Duplication:** De-Duplication ("De-Duping") is the process of comparing electronic records based on their characteristics and removing or marking duplicate records within the data set. The definition of "duplicate records" should be agreed upon, i.e., whether an exact copy from a different location (such as a different mailbox, server tapes, etc.) is considered to be a duplicate. De-duplication can be selective, depending on the agreed-upon criteria. *See also* Case De-Duplication, Content Comparison, Cross-Custodian De-Duplication, Custodian De-Duplication, Data Verification, Digital Fingerprint, File Level Binary Comparison, Hash Coding, Horizontal De-Duplication, Metadata Comparison, and Production De-Duplication.

*De-Fragment ("de-frag"):* Use of a computer utility to reorganize files so they are more contiguous on a hard drive or other storage medium, if the files or parts thereof have become fragmented and scattered in various locations within the storage medium in the course of normal computer operations. Used to optimize the operation of the computer, it will overwrite information in unallocated space. *See* Fragmented.

**Deleted Data:** Deleted Data is data that existed on the computer as live data and which have been deleted by the computer system or end-user activity. Deleted data may remain on storage media in whole or in part until they are overwritten or "wiped." Even after the data itself have been wiped, directory entries, pointers or other information relating to the deleted data may remain on the computer. "Soft deletions" are data marked as deleted (and not generally available to the end-user after such marking), but not yet physically removed or overwritten. Soft-deleted data can be restored with complete integrity.

*Deleted File:* A file with disc space that has been designated as available for reuse; the deleted file remains intact until it is overwritten.

*Deletion:* Deletion is the process whereby data is removed from active files and other data storage structures on computers and rendered inaccessible except through the use of special data recovery tools

designed to recover deleted data. Deletion occurs on several levels in modern computer systems: (a) *File level deletion* renders the file inaccessible to the operating system and normal application programs and marks the storage space occupied by the file's directory entry and contents as free and available to re-use for data storage, (b) *Record level deletion* occurs when a record is rendered inaccessible to a database management system (DBMS) (usually marking the record storage space as available for re-use by the DBMS, although in some cases the space is never reused until the database is compacted) and is also characteristic of many email systems (c) *Byte level deletion* occurs when text or other information is deleted from the file content (such as the deletion of text from a word processing file); such deletion may render the deleted data inaccessible to the application intended to be used in processing the file, but may not actually remove the data from the file's content until a process such as compaction or rewriting of the file causes the deleted data to be overwritten.

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

*De-shading*: Removing shaded areas to render images more easily recognizable by OCR. De-shading software typically searches for areas with a regular pattern of tiny dots.

*De-skewing*: The process of straightening skewed (tilted) images. De-skewing is one of the image enhancements that can improve OCR accuracy. Documents often become skewed when scanned or faxed.

*Desktop*: Generally refers to an individual PC.

*De-speckling*: Removing isolated speckles from an image file. Speckles often develop when a document is scanned or faxed.

*DIA/DCA (Document Interchange Architecture)*: An IBM standard for transmission and storage of voice, text or video over networks.

*Digital Certificate*: Electronic records that contain keys used to decrypt information, especially information sent over a public network like the Internet.

*Digital Fingerprint*: A fixed-length hash code that uniquely represents the binary content of a file. *See also* Data Verification *and* File Level Binary Comparison *and* Hash Coding.

*Digital*: Information stored as a string of ones and zeros. Opposite of analog.

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

*Directory*: A simulated file folder or container used to organize files and directories in a hierarchical or tree-like structure. UNIX and DOS use the term "directory", while Mac and Windows use the term "folder."

**Disaster Recovery Tapes**: Portable media used to store data for backup purposes. *See* Backup Data/Backup Tapes.

*Disc mirroring*: A method of protecting data from a catastrophic hard disc failure or for long term data storage. As each file is stored on the hard disc, a "mirror" copy is made on a second hard disc or on a different part of the same disc. *See also* Mirror.

*Disc Partition*: A hard drive containing a set of consecutive cylinders.



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

*Discovery:* Discovery is the process of identifying, locating, securing and producing information and materials for the purpose of obtaining evidence for utilization in the legal process. The term is also used to describe the process of reviewing all materials which may be potentially relevant to the issues at hand and/or which may need to be disclosed to other parties, and of evaluating evidence to prove or disprove facts, theories or allegations. There are several ways to conduct discovery, the most common of which are interrogatories, requests for production of documents and depositions.

*Discwipe:* Utility that overwrites existing data. Various utilities exist with varying degrees of efficiency - some wipe only named files or unallocated space of residual data, thus unsophisticated users who try to wipe evidence may leave behind files of which they are unaware.

**Disposition:** The final business action carried out on a record. This action generally is to destroy or archive the record. Electronic record disposition can include “soft deletions” (*see* Deletion), “hard deletions,” “hard deletions with overwrites,” “archive to long-term store,” “forward to organization,” and “copy to another media or format and delete (hard or soft).”

**Distributed Data:** Distributed Data is that information belonging to an organization which resides on portable media and non-local devices such as remote offices, home computers, laptop computers, personal digital assistants (“PDAs”), wireless communication devices (*e.g.*, Blackberry) and internet repositories (including email hosted by internet service providers or portals and web sites). Distributed data also includes data held by third parties such as application service providers and business partners. *Note:* Information Technology organizations may define distributed data differently (for example, in some organizations distributed data includes any non-server-based data, including workstation disc drives).

*Dithering:* In printing, dithering is usually called *half-toning*, and shades of gray are called *halftones*. The more dither patterns that a device or program supports, the more shades of gray it can represent. Dithering is the process of converting grays to different densities of black dots, usually for the purposes of printing or storing color or grayscale images as black and white images.

*DLT (Digital Linear Tape):* A type of backup tape which can hold up to 80 GB depending on the data file format.

*Document:* A page, a collection of pages or any file produced manually or by a software application, that constitutes a logical single communication of information. Examples include a letter, a spreadsheet or an email.

*Document Date:* The original creation date of a document. For an email the document date is indicated by the date-stamp of the email.

*Document Imaging Programs:* Software used to store, manage, retrieve and distribute documents quickly and easily on the computer.

*Document Metadata:* Data about the document stored in the document, as opposed to document content. Often this data is not immediately viewable in the software application used to create/edit the document but often can be accessed via a “Properties” view. Examples include document author and company, and create and revision dates. Contrast with File System Metadata and Email Metadata. *See also* Metadata.

*Document Type or Doc Type:* A typical field used in bibliographical coding. Typical doc type examples include letter, memo, report, article and others.

*Domain:* A sub-network of servers and computers within a LAN. Domain information is useful when restoring backup tapes, particularly of email.

*Domino Database:* Another name for Lotus Notes Databases versions 5.0 or higher. See NSF.

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

*DPI (Dots Per Inch):* The measurement of the resolution of display in printing systems. A typical CRT screen provides 96 dpi, which provides 9,216 dots per square inch (96x96). When a paper document is scanned, the resolution, or level of detail, at which the scanning was performed is expressed in DPI. Typically, documents are scanned at 200 or 300 DPI.

*Draft Record:* A draft record is a preliminary version of a record before it has been completed, finalized, accepted, validated or filed. Such records include working files and notes. Records and information management policies may provide for the destruction of draft records upon finalization, acceptance, validation or filing of the final or official version of the record. However, draft records generally must be retained if (1) they are deemed to be subject to a legal hold; or (2) a specific law or regulation mandates their retention and policies should recognize such exceptions.

*Drag-and-Drop:* The movement of on-screen objects by dragging them with the mouse, and dropping them in another place.

*DRAM:* Dynamic Random Access Memory, a memory technology which is periodically “refreshed” or updated – as opposed to “static” RAM chips which do not require refreshing. The term is often used to refer to the memory chips themselves.

*Drive Geometry:* A computer hard drive is made up of a number of rapidly rotating platters that have a set of read/write heads on both sides of each platter. Each platter is divided into a series of concentric rings called tracks. Each track is further divided into sections called sectors, and each sector is subdivided into bytes. Drive geometry refers to the number and positions of each of these structures.

*Driver:* A driver is a computer program that controls various devices such as the keyboard, mouse, monitor, etc.

*DSP (Digital Signal Processor/Processing):* A special purpose computer (or technique) which digitally processes signals and electrical/analog waveforms.

*DTP (Desktop Publishing):* PC applications used to prepare direct print output or output suitable for printing presses.

*Duplex Scanners vs. Double-Sided Scanning:* Duplex scanners automatically scan both sides of a double-sided page, producing two images at once. Double-sided scanning uses a single-sided scanner to scan double-sided pages, scanning one collated stack of paper, then flipping it over and scanning the other side.

*Duplex:* Two-sided page(s).

*DVD (Digital Video Disc or Digital Versatile Disc):* A plastic disc, like a CD, on which data can be written and read. DVDs are faster, can hold more information, and can support more data formats than CDs.

*ECM:* Enterprise content management.

*EDI (Electronic Data Interchange):* Eliminating forms altogether by encoding the data as close as possible to the point of the transaction; automated business information exchange.

*EDMS (Electronic Document Management System):* A system to electronically manage documents during all life cycles. *See* Electronic Document Management.

*EGA (Extended Graphics Adapter):* *See* VGA.

*EIA:* Electronic Industries Association *EIM:* Electronic Image Management.

*EISA (Extended Industry Standard Architecture):* One of the standard buses used for PCs.

*Electronic Discovery:* The process of collecting, preparing, reviewing, and producing electronic documents in the context of the legal process. *See* Discovery.

*Electronic Document Management:* For paper documents, involves imaging, indexing/coding and archiving of scanned documents/images, and thereafter electronically managing them during all life cycle phases. Electronic documents are likewise electronically managed from creation to archiving and all stages in between.

*Electronic File Processing:* Generally includes extraction of metadata from files, identification of duplicates/de-duplication and rendering of data into delimited format.

*Electronic Image:* An electronic or digital picture of a document (e.g. TIFF, PDF, etc.).

**Electronic Mail Message:** A document created or received via an electronic mail system, including brief notes, formal or substantive narrative documents, and any attachments, such as word processing and other electronic documents, which may be transmitted with the message.

*Electronic Mail/Email:* An electronic means for communicating information under specified conditions, generally in the form of text messages, through systems that will send, store, process, and receive information and in which messages are held in storage until the addressee accesses them.

**Electronic Record:** Information recorded in a form that requires a computer or other machine to process it and that otherwise satisfies the definition of a record.

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

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

*Email address:* An electronic mail address. Internet email addresses follow the formula: user-ID@domain-name; other email protocols may use different address formats. In some email systems, a user’s email address is “aliased” or represented by his or her natural name rather than a fully qualified email address. For example, john.doe@abc.com might appear simply as John Doe.

*Email Metadata:* Data stored in the email about the email. Often this data is not even viewable in the email client application used to create the email. The amount of email metadata available for a particular email varies greatly depending on the email system. Contrast with File System Metadata and Document Metadata.

*Email String:* A series of emails linked together by email responses or forwards. The series of email messages created through multiple responses and answers to an originating message. Also referred to as an email "thread." Comments, revisions, attachments are all part of an email string.

*Encryption:* A procedure that renders the contents of a message or file scrambled or unintelligible to anyone not authorized to read it. Encryption is used to protect information as it moves from one computer to another and is an increasingly common way of sending credit card numbers and other personal information over the Internet.

*Encryption Key:* A data value that is used to encrypt and decrypt data. The number of bits in the encryption key is a rough measure of the encryption strength; generally, the more bits in the encryption key, the more difficult it is to break.

*End Document Number or End Doc#:* The last single page image of a document.

*Endorser:* A small printer in a scanner that adds a document-control number or other endorsement to each scanned sheet.

*Enhanced Titles:* A meaningful/descriptive title for a document. The opposite of Verbatim Titles.

*Enterprise Architecture:* Framework for how software, computing, storage and networking systems should integrate and operate to meet the changing needs across an entire business

*EOF (End of File):* A distinctive code which uniquely marks the end of a data file.

*EPP (Enhanced Parallel Port):* Also known as Fast Mode Parallel Port. A new, industry standard parallel port, having higher transfer times competitive with SCSI.

*EPS (Encapsulated PostScript):* Uncompressed files for images, text and objects. Only print on PostScript printers.

*Erasable Optical Drive:* A type of optical drive that uses erasable optical discs.

*ESDI (Enhanced Small Device Interface):* A defined, common electronic interface for transferring data between computers and peripherals, particularly disc drives.

*ESI:* Electronically stored information.

*Ethernet:* A common way of networking PCs to create a Local Area Network (LAN).

*Evidentiary Image or Copy:* See Forensic Copy.

*Exabyte:* A unit of 1000 petabytes. See Byte.

*Export:* Data extracted or taken out of one environment or application usually in a prescribed format, and usually for import into another environment or application.

*Extended Partitions:* If a computer hard drive has been divided into more than four partitions, extended partitions are created. Under such circumstances each extended partition contains a partition table in the first sector that describes how it is further subdivided.

*Extensible Markup Language (XML):* Short for Extensible Markup Language, a specification developed by the W3C (World Wide Web Consortium—the Web development standards board). XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tag, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

*Extranet:* An Internet based access method to a corporate intranet site by limited or total access through a security firewall. This type of access is often utilized in cases of joint defense, joint venture and vendor client relationships.

*False Positive/Negative:* A result that is not correct. This may be a result of performing a process incorrectly or using a process that is not accurate.

*FAT (File Allocation Table):* An internal data table on hard drives that keeps track of where the files are stored. If a FAT is corrupt, a drive may be unusable, yet the data may be retrievable with forensics. *See* Cluster File.

*FAX:* Short for facsimile. A process of transmitting documents by scanning them to digital, converting to analog, transmitting over phone lines and reversing the process at the other end and printing.

*Fiber Optics:* Transmitting information by sending light pulses over cables made from thin strands of glass.

*Field (or Data Field):* A name for an individual piece of standardized data, such as the author of a document, a recipient, the date of a document or any other piece of data common to most documents in an image collection, to be extracted from the collection.

*Field Separator:* A code that separates the fields in a record. For example, the CSV format uses a comma as the field separator.

*File Compression:* *See* Compression.

*File Extension:* Many systems, including DOS and UNIX, allow a filename extension that consists of one or more characters following the proper filename. For example, image files are usually stored as .bmp, .gif, .jpg or .tiff. Audio files are often stored as .aud or .wav. There are a multitude of file extensions identifying file formats. The filename extension should indicate what type of file it is; however, users may change filename extensions to evade firewall restrictions or for other reasons. Therefore, file types should be identified at a binary level rather than relying on file extensions. To research file types, see (<http://www.filext.com>). Different applications can often recognize only a predetermined selection of file types. *See also* Format.

*File Format:* The organization or characteristics of a file that determine with which software programs it can be used. *See also* Format.

*File Level Binary Comparison:* Method of de-duplication using the digital fingerprint (hash) of a file. File Level Binary comparison ignores metadata, and can determine that “SHOPPING LIST.DOC” and

“TOP SECRET.DOC” are actually the same document. *See* Data Verification, Digital Fingerprint, and Hash coding. *See* De-Duplication.

**File Plan:** A document containing the identifying number, title, description, and disposition authority of files held or used in an office.

**File Server:** When several or many computers are networked together in a LAN situation, one computer may be utilized as a storage location for files for the group. File servers may be employed to store email, financial data, word processing information or to back-up the network. *See* Server.

**File Sharing:** Sharing files stored on the server among several users on a network.

**File Slack:** The unused space on a cluster that exists when the logical file space is less than the physical file space.

**File System Metadata:** Data that can be obtained or extracted about a file from the file system storing the file. Examples include file creation time, last modification time, and last access time.

**File System:** The engine that an operating system or program uses to organize and kept track of files. More specifically, the logical structures and software routines used to control access to the storage on a hard disc system and the overall structure in which the files are named, stored, and organized. The file system plays a critical role in computer forensics because the file system determines the logical structure of the hard drive, including its cluster size. The file system also determines what happens to data when the user deletes a file or subdirectory.

**File Transfer:** The process of moving or transmitting a file from one location to another, as between two programs or from one computer to another.

**File:** A collection of data or information stored under a specified name on a disc.

**Filename:** The name of a file, excluding root drive and directory path information. Different operating systems may impose different restrictions on filenames, for example, by prohibiting use of certain characters in a filename or imposing a limit on the length of a filename. The filename extension should indicate what type of file it is. However, users often change filename extensions to evade firewall restrictions or for other reasons. Therefore, file types must be identified at a binary level rather than relying on file extensions. *See also* File Extension *and* Full Path.

**FIPS:** Federal Information Processing Standards issued by the National Institute of Standards and Technology after approval by the Secretary of Commerce pursuant to Section 111(d) of the Federal Property and Administrative Services Act of 1949, as amended by the Computer Security Act of 1987, Public Law 100-235.

**Firewall:** A set of related programs, or hardware, that protect the resources of a private network from users from other networks. A firewall filters information to determine whether to forward the information toward its destination.

**Filter (verb):** *See* Data Filtering.

**Flatbed Scanner:** A flat-surface scanner that allows users to input books and other documents.

*Floppy Disc:* A thin magnetic film disc housed in a protective sleeve used to copy and transport relatively small amounts of data.

*Folder:* See Directory.

**Forensic Copy:** A forensic copy is an exact copy of an entire physical storage media (hard drive, CD-ROM, DVD-ROM, tape, etc.), including all active and residual data and unallocated space on the media. Compresses and encrypts to ensure authentication and protect chain of custody. Forensic copies are often called “image or imaged copies.” See Bit Stream Back-up and Mirroring.

*Forensics:* Computer forensics is the scientific examination and analysis of data held on, or retrieved from, computer storage media in such a way that the information can be used as evidence in a court of law. It may include the secure collection of computer data; the examination of suspect data to determine details such as origin and content; the presentation of computer based information to courts of law; and the application of a country’s laws to computer practice. Forensics may involve recreating “deleted” or missing files from hard drives, validating dates and logged in authors/editors of documents, and certifying key elements of documents and/or hardware for legal purposes.

*Form of Production:* The manner in which requested documents are produced. Used to refer both to file format (native vs. PDF or TIFF) and the media on which the documents are produced (paper vs. electronic).

**Format (noun):** The internal structure of a file, which defines the way it is stored and used. Specific applications may define unique formats for their data (e.g., “MS Word document file format”). Many files may only be viewed or printed using their originating application or an application designed to work with compatible formats. There are several common email formats, such as Outlook and Lotus Notes. Computer storage systems commonly identify files by a naming convention that denotes the format (and therefore the probable originating application). For example, “DOC” for Microsoft Word document files; “XLS” for Microsoft Excel spreadsheet files; “TXT” for text files; “HTM” for Hypertext Markup Language (HTML) files such as web pages; “PPT” for Microsoft Powerpoint files; “TIF” for tiff images; “PDF” for Adobe images; etc. Users may choose alternate naming conventions, but this will likely affect how the files are treated by applications.

*Format (verb):* Makes a drive ready for first use. Erroneously thought to “wipe” drive. Typically, only overwrites FAT, but not files on the drive.

*Forms Processing:* A specialized imaging application designed for handling pre-printed forms. Forms processing systems often use high-end (or multiple) OCR engines and elaborate data validation routines to extract hand-written or poor quality print from forms that go into a database.

*Fragmented:* In the course of normal computer operations when files are saved, deleted or moved, the files or parts thereof may be broken into pieces, or fragmented, and scattered in various locations on the computer’s hard drive or other storage medium, such as removable discs. Data saved in contiguous clusters may be larger than contiguous free space, and it is broken up and randomly placed throughout the available storage space. See De-Fragment.

*FTP (File Transfer Protocol):* An Internet protocol that enables the transfer of files between computers over a network or the Internet.

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

*Full Path:* A path name description that includes the drive, starting or root directory, all attached subdirectories and ending with the file or object name.

*Full-Text Search:* The ability to search a data file for specific words, numbers and/or combinations or patterns thereof.

*Full-Text Indexing and Search:* Every word in the document is indexed into a master word list with pointers to the documents and pages where each occurrence of the word appears.

*Fuzzy Search:* Subjective content searching (as compared to word searching of objective data). Fuzzy Searching lets the user find documents where word matching does not have to be exact, even if the words searched are misspelled due to optical character recognition (OCR) errors.

*GAL:* A Microsoft Outlook global address list - directory of all Microsoft Exchange users and distribution lists to whom messages can be addressed. The administrator creates and maintains this list. The global address list may also contain public folder names. Entries from this list can be added to a user's personal address book.

*Ghost:* See Bit Stream Back-up.

*GIF(Graphics Interchange Format):* CompuServe's native file format for storing images. Limited to 256 colors.

*Gigabyte (GB):* A unit of consisting of either 1,000 or 1,024 megabytes. In terms of image storage capacity, one gigabyte equals approximately 17,000 8 1/2" x 11" pages scanned at 300 dpi, stored as TIFF Group IV images. See Byte.

*GMT Timestamp:* Identification of a file using Greenwich Mean Time as the central time authentication method.

*GPS Generated Timestamp:* Timestamp identifying time as a function of its relationship to Greenwich Mean Time.

*Gray Scale:* The use of many shades of gray to represent an image. *Continuous-tone* images, such as black-and-white photographs, use an almost unlimited number of shades of gray. Conventional computer hardware and software, however, can only represent a limited number of shades of gray (typically 16 or 256).

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

*GUI (Graphical User Interface, pronounced "gooey"):* Presenting an interface to the computer user comprised of pictures and icons, rather than words and numbers.

*Hacker:* Someone who breaks into computer systems in order to steal, change or destroy information.

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

*Halftone:* See Dithering.



*Hard Disc Drive:* The primary storage unit on PCs, consisting of one or more magnetic media platters on which digital data can be written and erased magnetically.

*Harvesting:* The process of retrieving or collecting electronic data from storage media or devices; an E-Discovery vendor “harvests” electronic data from computer hard drives, file servers, CDs, and backup tapes for processing and load to storage media or a database management system.

*Hash:* A mathematical algorithm that represents a unique value for a given set of data, similar to a digital fingerprint. Common hash algorithms include MD5 and SHA.

*Hash Coding:* To create a digital fingerprint that represents the binary content of a file unique to every electronically-generated document; assists in subsequently ensuring that data has not been modified. *See also Data Verification and Digital Fingerprint and File Level Binary Comparison.*

*Hash Function:* A function used to create a hash value from binary input. The hash is substantially smaller than the text itself, and is generated by the hash function in such a way that it is extremely unlikely that some other input will produce the same hash value.

*HD (High Density):* A 5.25” HD Floppy Disc holds 1.2 MB and a 3.5” holds 1.4 MB.

*Head:* Each platter on a hard drive contains a head for each side of the platter. The heads are devices which ride very closely to the surface of the platter and allow information to be read from and written to the platter.

*Hexadecimal:* A number system with a base of 16. The digits are 0-9 and A-F, where F equals the decimal value of 15.

*Hidden Files or Data:* Files or data not visible in the file directory; cannot be accessed by unauthorized or unsophisticated users. Some operating system files are hidden, to prevent inexperienced users from inadvertently deleting or changing these essential files. *See also Steganography.*

*Hierarchical Storage Management (HSM):* Software that automatically migrates files from on-line to near-line storage media, usually on the basis of the age or frequency of use of the files.

*Hold:* *See Legal Hold.*

*Holorith:* encoded data on aperture cards *or* old-style punch cards that contained encoded data.

*Horizontal De-duplication:* A way to identify documents that are duplicated across multiple custodians or other production data sets. *See De-Duplication.*

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

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

*HTML:* HyperText Markup Language, developed by CERN of Geneva, Switzerland. The document standard of choice of Internet. (HTML+ adds support for multi-media.) The tag-based ASCII language used to create pages on the World Wide Web - uses tags to tell a web browser to display text and images.

*HTTP (HyperText Transfer Protocol):* The underlying protocol used by the World Wide Web. HTTP defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. For example, when you enter a URL in your browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page.

*Hub:* A network device that connects multiple computers/peripherals together and allows them to share data. A central unit that repeats and/or amplifies data signals being sent across a network.

*Hyperlink:* A link - usually appearing as a highlighted word or picture within a hypertext document - that when clicked changes the active view, possibly to another place within the same document or view, or to another document altogether, usually regardless of the application or environment in which the other document or view exists.

*HyperText:* Text that includes links or shortcuts to other documents or views, allowing the reader to easily jump from one view to a related view in a non-linear fashion.

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

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

*IDE (Integrated Drive Electronics):* An engineering standard for interfacing PC’s and hard discs.

*IEEE (Institute of Electrical and Electronic Engineers):* An international association which sponsors meetings, publishes a number of journals and establishes standards.

*ILM:* Information lifecycle management.

*Image:* To image a hard drive is to make an identical copy of the hard drive, including empty sectors. Also known as creating a “mirror image” or “mirroring” the drive.

**Image Copy, Imaged Copy:** *See* Forensic Copy.

*Image Enabling:* A software function that creates links between existing applications and stored images.

*Image File Format:* *See* File Format and Format.

*Image Key:* The name of a file created when a page is scanned in a collection.

*Image Processing Card (IPC):* A board mounted in the computer, scanner or printer that facilitates the acquisition and display of images. The primary function of most IPCs is the rapid compression and decompression of image files.

*Image Processing:* To capture an image or representation, usually from electronic data in native format, enter it in a computer system, and process and manipulate it. *See also* Native Format.

*Import:* Data brought into an environment or application which has been exported from another environment or application.

**Inactive Record:** Inactive records are those Records related to closed, completed, or concluded activities. Inactive Records are no longer routinely referenced, but must be retained in order to fulfill reporting requirements or for purposes of audit or analysis. Inactive records generally reside in a long-term storage format remaining accessible for purposes of business processing only with restrictions on alteration. In some business circumstances inactive records may be re-activated.

*Index/Coding Fields:* Database fields used to categorize and organize documents. Often user-defined, these fields can be used for searches.

*Index:* The searchable catalog of documents created by search engine software. Also called “catalog.” Index is often used as a synonym for search engine.

*Indexing:* Universal term for Coding and Data Entry.

**Information:** For the purposes of this document, information is used to mean both documents and data.

*Input device:* Any peripheral that allows a user to communicate with a computer by entering information or issuing commands (e.g., keyboard).

**Instant Messaging (“IM”):** A form of electronic communication involving immediate correspondence between two or more online users. Peer-to-peer IM communications may not be stored on servers after receipt; logging of peer-to-peer IM messages is typically done on the client computer, and may be optionally enabled or disabled on each client.

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

*Interleave:* To arrange data in a noncontiguous way to increase performance. When used to describe disc drives, it refers to the way sectors on a disc are organized. In one-to-one interleaving, the sectors are placed sequentially around each track. In two-to-one interleaving, sectors are staggered so that consecutively numbered sectors are separated by an intervening sector. The purpose of interleaving is to make the disc drive more efficient. The disc drive can access only one sector at a time, and the disc is constantly spinning beneath.

*International Telecommunication Union (ITU):* An international organization under the UN headquartered in Geneva concerned with telecommunications that develops international data communications standards; known as CCITT prior to March 1, 1993. See <http://www.itu.int>.

*Internet Publishing:* Specialized imaging software that allows documents to be published on the Internet

*Internet:* A worldwide network of networks that all use the TCP/IP communications protocol and share a common address space. It supports services such as email, the World Wide Web, file transfer, and Internet Relay Chat. Also known as “the net”, “the information superhighway”, and “cyberspace”.

*Inter-Partition Space:* Unused sectors on a track located between the start of the partition and the partition boot record. This space is important because it is possible for a user to hide information here.