

**IMPORTANT NOTICE:**  
**This Publication Has Been Superseded**

**See the Most Current Publication at**

[https://thesedonaconference.org/publication/The\\_Sedona\\_Conference\\_Glossary](https://thesedonaconference.org/publication/The_Sedona_Conference_Glossary)



# 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®.



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

# *The Sedona Conference® Glossary*

---

## 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 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 *halftoning*, 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 sub-divided 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 keep 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.

**Intranet:** A private network that uses Internet-related technologies to provide services within an organization.

**IP address (Internet Protocol address):** A string of four numbers separated by periods used to represent a computer on the Internet - a unique identifier for the physical location of the server containing the data. See TCP/IP (e.g., 206-1432.001).

**IPX/SPX:** Communications protocol used by Novell networks.

**IS/IT Information Systems or Information Technology:** Usually refers to the people who make computers and computer systems run.

**ISA:** Industry Standard Architecture.

**ISDN (Integrated Services Digital Network):** An all digital network which can carry data, video and voice.

**ISIS and TWAIN Scanner Drivers:** Specialized applications used for communication between scanners and computers.

**ISO 9660 CD Format:** The International Standards Organization format for creating CD-ROMs that can be read worldwide.

**ISO:** International Standards Organization.

**ISP (Internet Service Provider):** A business that provides access to the Internet, usually for a monthly fee. ISPs may be a source of evidence through files (such as ISP email) stored on ISP servers.

**IT (Information Technology) Infrastructure:** The overall makeup of business-wide technology operations, including mainframe operations, standalone systems, email, networks (WAN and LAN), internet access, customer databases, enterprise systems, application support, regardless of whether managed, utilized or provided locally, regionally, globally, etc., or whether performed or located internally or by outside providers (outsourced to vendors). The IT Infrastructure also includes applicable standard practices and procedures, such as backup procedures, versioning, resource sharing, retention practices, janitor program utilization, and the like.

**Janitor Program:** An application which runs at scheduled intervals to manage business information by deleting, transferring, or archiving on-line data (such as email) which is at or past its scheduled active life. Janitor programs are sometimes referred to as "agents"—software that runs autonomously "behind the scenes" on user systems and servers to carry out business processes according to pre-defined rules. Janitor programs must include a facility to support disposition and process holds.

**Java:** Sun Microsystems' Java is a platform-independent, programming language for adding animation and other actions to websites.

**Jaz Drive:** A removable disc drive. A Jaz drive holds up to 2 GB of data. Commonly used for backup storage as well as everyday use.

**JMS:** Jukebox Management Software.

*Journal:* A chronological record of data processing operations that may be used to reconstruct a previous or an updated version of a file. In database management systems, it is the record of all stored data items that have values changed as a result of processing and manipulation of the data.

*Journaling:* A function of e-mail systems (such as Microsoft Exchange and Lotus Notes) that copies sent and received items into a second information store for retention or preservation. Because Journaling takes place at the information store (server) level when the items are sent or received, rather than at the mailbox (client) level, some message-related metadata, such as user foldering (what folder the item is stored in within the recipient's mailbox) and the status of the "read" flag, is not retained in the journaled copy. The Journaling function stores items in the system's native format, unlike e-mail archiving solutions, which use proprietary storage formats that are designed to reduce the amount of storage space required. Journaling systems also lack the sophisticated search and retrieval capabilities contained in e-mail archiving solutions.

*JPEG (Joint Photographic Experts Group):* A compression algorithm for still images that is commonly used on the web.

*Jukebox:* A mass storage device that holds optical disc s and loads them into a drive.

*Juke-Box:* Automated disc changer for high-performance, centralized storage for multifunction CD-ROM's & optical discs.

*Jump Drive:* See Key Drive.

*Kerning:* Adjusting the spacing between two letters.

*Key Drive:* A small removable data storage device that uses flash memory and connects via a USB port. Keydrives are also known as keychain drive, thumb drive, jump drive, USB flash drive. Can be imaged and may contain residual data.

*Key Field:* Database fields used for document searches and retrieval.

*Keyword Search:* A search of the text of documents in a database for documents containing one or more words that are specified by a user.

*Keywords:* Words designated by a user as important for searching purposes.

*Kilobyte (KB):* A unit of 1,024 bytes. See Byte.

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

*LAN (Local Area Network):* A group of computers at a single location (usually an office or home) that are connected by phone lines or coaxial cable. See Network.

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

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

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

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

**Latent Data, Legacy System:** Latent or ambient data are deleted files and other data that are inaccessible without specialized forensic tools and techniques. Until overwritten, these data reside on media such as a hard drive in unused space and other areas available for data storage.

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

**Legacy Data, Legacy System:** Legacy Data is information the development of which an organization may have invested significant resources and has retained its importance, but has been created or stored by the use of software and/or hardware that has become obsolete or replaced (“legacy systems”). Legacy data may be costly to restore or reconstruct when required for investigation or litigation analysis or discovery.

**Legal Hold:** A legal hold is a communication issued as a result of current or anticipated litigation, audit, government investigation or other such matter that suspends the normal disposition or processing of records. Legal holds can encompass business procedures affecting active data, including, but not limited to, backup tape recycling. The specific communication to business or IT organizations may also be called a “hold,” “preservation order,” “suspension order,” “freeze notice,” “hold order,” or “hold notice.”

**Level Coding:** Used in Bibliographical coding to facilitate different treatment, such as prioritization or more thorough extraction of data, for different categories of documents, such as by type or source.

**LFP:** IPRO Tech’s image cross reference file; an ASCII delimited text file required for cross-reference of images to data.

**Lifecycle:** The records lifecycle is the life span of a record from its creation or receipt to its final disposition. It is usually described in three stages: creation, maintenance and use, and archive to final disposition.

**Line Screen:** The number of half-tone dots that can be printed per inch. As a general rule, newspapers print at 65 to 85 lpi.

**Link:** See Hyperlink.

**Load file:** A file that relates to a set of scanned images and indicates where individual pages belong together as documents. A load file may also contain data relevant to the individual documents, such as metadata, coded data and the like. Load files must be obtained and provided in prearranged formats to ensure transfer of accurate and usable images and data.

**Local Area Network (LAN):** See Network.

**Logical File Space:** The actual amount of space occupied by a file on a hard drive. The amount of logical file space differs from the physical file space because when a file is created on a computer, a sufficient number of clusters (physical file space) are assigned to contain the file. If the file (logical file space) is

not large enough to completely fill the assigned clusters (physical file space) then some unused space will exist within the physical file space.

*Logical Volume:* An area on the hard drive that has been formatted for files storage. A hard drive may contain a single or multiple volumes.

*Lossless Compression:* Exact construction of image, bit-by-bit, with no loss of information.

*Lossy Compression:* Reduces storage size of image by reducing the resolution and color fidelity while maintaining minimum acceptable standard for general use.

*LTO (Linear Tape-Open):* A type of backup tape which can hold as much as 400 GB of data, or 600 CDs depending on the data file format.

*LZW (Lempel-Ziv & Welch):* A common, lossless compression standard for computer graphics, used for most TIFF files. Typical compression ratios are 4/1.

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

*Magnetic/Optical Storage Media:* Includes, but is not limited to, hard drives, backup tapes, CD-ROMs, DVD-ROMs, Jaz and Zip drives.

*Magneto-Optical Drive:* A drive that combines laser and magnetic technology to create high-capacity erasable storage.

*Mailbox:* An area on a storage device where email is placed. In email systems, each user has a private mailbox. When the user receives email, the mail system automatically puts it in the appropriate mailbox.

*Make-Available Production:* A process whereby what is usually a large universe of all potentially responsive documents are made available to the requestor; from this universe, the requestor then reviews and selects or tags the documents which they wish to obtain, and the producing party produces to the requestor only the selected documents. This is sometimes done under an agreement protecting against privilege and confidentiality waiver during the initial make available production; and the producing party, after the requestor has selected the documents they wish to obtain, reviews only the selected documents for privilege and confidentiality before the selected documents are physically produced to the requestor.

*MAPI (Mail Application Program Interface):* A Windows software standard that has become a popular email interface used by MS Exchange, GroupWise, and other email packages.

*MAPI Mail Near-Line:* Documents stored on optical discs or compact discs that are housed in the jukebox or CD changer and can be retrieved without human intervention.

*Marginalia:* Handwritten notes in the margin of the page in documents.

*Master Boot Record:* See Boot Sector.

*Mastering:* Making many copies of a disc from a single master disc.

*MCA (Micro Channel Architecture):* An IBM bus standard.

**MDE (Magnetic Disc Emulation):** Software that makes a jukebox look and operate like a hard-drive such that it will respond to all the I/O commands ordinarily sent to a hard drive.

**MD5:** message-digest algorithm meant for digital signature applications where a large message has to be “compressed” in a secure manner before being signed with the private key

**Media:** An object or device, such as a disc, tape, or other device, on which data is stored

**Megabyte (M or MB):** A unit of approximately 1 million bytes or 1024 KB. *See* Byte.

**Memory:** Data storage in the form of chips, or the actual chips used to hold data; “storage” is used to describe memory that exists on tapes or discs. See RAM and ROM.

**Menu:** A list of options, each of which performs a desired action such as choosing a command or applying a particular format to a part of a document.

**Message Header:** Message headers generally contain the identities of the author and recipients, the subject of the message, and the date the message was sent.

**Metadata:** Metadata is information about a particular data set or document which describes how, when and by whom it was collected, created, accessed, modified and how it is formatted. Can be altered intentionally or inadvertently. Can be extracted when native files are converted to image. Some metadata, such as file dates and sizes, can easily be seen by users; other metadata can be hidden or embedded and unavailable to computer users who are not technically adept. Metadata is generally not reproduced in full form when a document is printed. *See also* Customer-Added Metadata, Document Metadata, Email Metadata, File System Metadata, User-Added Metadata and Vendor-Added Metadata. For a thorough discussion of Metadata, *see The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age, Appendix D: Technical Appendix.*

**Metadata Comparison:** A method of de-duplication that compares file metadata and ignores content. *See* De-Duplication.

**MFT (Master File Table):** Index to files on a computer. If corrupt, a drive may be unusable, yet data may be retrievable using forensic methods.

**MICR (Magnetic Ink Character Recognition):** The process used by banks to encode checks.

**Microfiche:** Sheet microfilm (4” by 6”) containing reduced images of 270 pages or more in a grid pattern.

**Migrated Data:** Migrated Data is information that has been moved from one database or format to another.

**Migration:** Moving files to another computer application or platform; may require conversion to a different format.

**Mirroring:** The duplication of data for purposes of backup or to distribute Internet or network traffic among several servers with identical data. *See also* Disc Mirroring.

**MIS:** Management Information Systems.

**MODEM:** Modulator/Demodulator. A device translates digital data from a computer into analog signals (modulates) and transmits the information over telephone lines. Another modem at the receiving computer will receive the information, translate it back from analog to digital (demodulate) and store it.

**Monochrome:** Displays capable of only two colors, usually black and white, or black and green.

**Mosaic:** A web browser popular before the introduction of Netscape and Internet Explorer.

**Mount, Mounting:** The process of making off-line data available for on-line processing. For example, placing a magnetic tape in a drive and setting up the software to recognize or read that tape. The terms “load” and “loading” are often used in conjunction with, or synonymously with, “mount” and “mounting” (as in “mount and load a tape”). “Load” may also refer to the process of transferring data from mounted media to another media or to an on-line system.

**MPEG-1 & -2:** Two different standards for full motion video to digital compression/decompression techniques advanced by the Moving Pictures Experts Group. MPEG-1 compresses 30 frames/second of full-motion video down to about 1.5 Mbits/sec from several hundred megabytes. MPEG-2 compresses the same files down to about 3.0 Mbits/sec and provides better image quality.

**MS-DOS:** Microsoft (MS)-Disc Operating System. Used in PCs as the control system.

**MTBF (Mean Time Between Failure):** Average time between failures. Used to compute the reliability of devices/equipment.

**MTTR (Mean Time To Repair):** Average time to repair. The higher the number, the more costly and difficult to fix.

**Multimedia:** The combined use of different media; integrated video, audio, text and data graphics in digital form.

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

**Native Format:** Electronic documents have an associated file structure defined by the original creating application. This file structure is referred to as the “native format” of the document. Because viewing or searching documents in the native format may require the original application (for example, viewing a Microsoft Word document may require the Microsoft Word application), documents are often converted to a vendor-neutral format as part of the record acquisition or archive process. “Static” formats (often called “imaged formats”), such as TIFF or PDF, are designed to retain an image of the document as it would look viewed in the original creating application but do not allow metadata to be viewed or the document information to be manipulated.

**Natural Language Search:** A manner of searching that permits the use of plain language without special connectors or precise terminology, such as “Where can I find information on William Shakespeare?” as opposed to formulating a search statement (such as “information” **and** “William Shakespeare”).

**Near-Line Data:** A term used to refer to data or a robotic storage device (robotic library) that houses removable media, uses robotic arms to access the media, and uses multiple read/write devices to store and retrieve records. Examples include optical discs.

**Near-Line Data Storage:** Storage in a system that is not a direct part of the network in daily use, but that can be accessed through the network. There is usually a small time lag between the request for data stored in near-line media and its being made available to an application or end-user. Making near-line data available will not require human intervention (as opposed to “off-line” data which can only be made available through human actions).

**Network Gear:** Refers to the actual hardware used in the operation of networks – for example routers, switches and hubs.

**Network:** A group of two or more computers and other devices connected together (“networked”) for the exchange and sharing of data and resources. A local-area network (LAN) refers to connected computers and devices geographically close together (i.e. in the same building). A wide-area network (WAN) refers generally to a network of PC’s or other devices, remote to each other, connected by telecommunications lines. Typically, a WAN may connect two or more LAN’s together.

**Neural Network:** Neural networks are made up of interconnected processing elements called units, which respond in parallel to a set of input signals given to each

**NIST - National Institute of Standards and Technology** - a federal technology agency that works with industry to develop and apply technology measurements and standards.

**Node:** Any device connected to a network. PCs, servers, and printers are all nodes on the network.

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

**NOS (Network Operating System):** See Operating System.

**NSF:** Lotus Notes Format Database File (i.e. database.nsf) Can be either an email database or the traditional type of fielded database.

**Objects:** In programming terminology, an object is a freestanding block of code that defines the properties of some thing. Objects are created and used in a high-level method of programming called object-oriented programming (OOP). OOP involves giving programming objects characteristics that can be transferred to, added to, and combined with other objects to make a complete program.

**OCR (Optical Character Recognition):** A technology process that translates and converts printed matter on an image into a format that a computer can manipulate (ASCII codes, for example) and, therefore, renders that matter text searchable. OCR software evaluates scanned data for shapes it recognizes as letters or numerals. All OCR systems include an optical scanner for reading text, and software for analyzing images. Most OCR systems use a combination of hardware (specialized circuit boards) and software to recognize characters, although some inexpensive systems operate entirely through software. Advanced OCR systems can read text in a large variety of fonts, but still have difficulty with handwritten text. OCR technology relies upon the quality of the imaged material, the conversion accuracy of the software, and the quality control process of the provider. The process is generally acknowledged to be only 80-85 percent accurate.

**Official Record Owner:** See Record Owner.

**Off-Line Data:** The storage of electronic data outside the network in daily use (e.g., on backup tapes) that is only accessible through the off-line storage system, not the network.

**Off-Line Storage:** Electronic records stored or archived on removable disc (optical, compact, etc.) or magnetic tape used for making disaster-recovery copies of records for which retrieval is unlikely. Accessibility to off-line media usually requires manual intervention and is much slower than on-line or near-line storage depending on the storage facility. The major difference between near-line data and offline data is that offline data lacks an intelligent disc subsystem, and is not connected to a computer, network, or any other readily-accessible system.

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

**On-Line Review:** The culling process produces a dataset of potentially responsive documents which are then reviewed for a final selection of relevant or responsive documents and assertion of privilege exception as appropriate. On-line Review enables the culled dataset to be accessed via PC or other terminal device via a local network or remotely via the Internet. Often, the On-Line Review process is facilitated by specialized software which provides additional features and functions which may include: collaborative access of multiple reviewers, security, user logging, search and retrieval, document coding, redaction, and privilege logging.

**On-Line Storage:** The storage of electronic data as fully accessible information in daily use on the network or elsewhere.

**Online/On-Line:** Connected (to a network).

**Operating System (OS):** An Operating system provides the software platform which directs the overall activity of a computer, network or system, and on which all other software programs and applications can run. In many ways, choice of an operating system will effect which applications can be run. Operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files and directories on the disc and controlling peripheral devices such as disc drives and printers. For large systems, the operating system has even greater responsibilities and powers - becoming a traffic cop to makes sure different programs and users running at the same time do not interfere with each other. The operating system is also responsible for security, ensuring that unauthorized users do not access the system. Examples of operating systems are UNIX, DOS, Windows, LINUX, Macintosh, and IBM's VM. Operating systems can be classified in a number of ways, including: multi-user (allows two or more users to run programs at the same time - some operating systems permit hundreds or even thousands of concurrent users); multiprocessing (supports running a program on more than one CPU); multitasking (allows more than one program to run concurrently); multithreading (allows different parts of a single program to run concurrently); and real time (instantly responds to input - general-purpose operating systems, such as DOS and UNIX, are not real-time).

**Optical Discs:** Computer media similar to a compact disc that cannot be rewritten. An optical drive uses a laser to read the stored data.

**Optical Jukebox:** See "Jukebox."

**OST:** A Microsoft Outlook information store that is used to save folder information that can be accessed offline.

**Overwrite:** To record or copy new data over existing data, as in when a file or directory is updated. Data that is overwritten cannot be retrieved.

**PAB:** A Microsoft Outlook list of recipients created and maintained by an individual user for personal use. The personal address book is a subset of the global address list (GAL).

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

**Packet:** A unit of data sent across a network which may contain identify and routing information. When a large block of data is to be sent over a network, it is broken up into several packets, sent, and then reassembled at the other end. The exact layout of an individual packet is determined by the protocol being used.

**Page File/Paging File:** A file used to temporarily store code and data for programs that are currently running. This information is left in the swap file after the programs are terminated, and may be retrieved using forensic techniques. Also referred to as a swap file.

**Page:** A single image of the equivalent of “one piece of paper”. One or several pages make up a “Document.”

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

**Partition:** A partition is an individual section of computer storage media such as a hard drive. For example a single hard drive may be divided into several partitions. When a hard drive is divided into partitions, each partition is designated by a separate drive letter, i.e., C, D, etc.

**Partition Table:** The partition table indicates each logical volume contained on a disc and its location.

**Partition Waste Space:** After the boot sector of each volume or partition is written to a track, it is customary for the system to skip the rest of that track and begin the actual useable area of the volume on the next track. This results in unused or “wasted” space on that track where information can be hidden. This “wasted space” can only be viewed with a low level disc viewer. However, forensic techniques can be used to search these “wasted space” areas for hidden information.

**Password:** A secret code utilized, usually along with a user ID, in order to log on or gain access to a PC, network or other secure system, site or application.

**Path:** The hierarchical description of where a directory, folder, or file is located on a computer or network. In DOS and Windows systems, a path is a list of directories where the operating system looks for executable files if it is unable to find the file in the working directory. The list of directories can be specified with the PATH command. Path is also used to refer to a transmission channel, the path between two nodes of a network that a data communication follows, and the physical cabling that connects the nodes on a network.

**Pattern Matching:** A generic term that describes any process that compares one file’s content with another file’s content.

**Pattern Recognition:** Technology that searches data for like patterns and flags, and extracts the pertinent data, usually utilizing an algorithm. For instance, in looking for addresses, alpha characters followed by a comma and a space, followed by two capital alpha characters, followed by a space, followed by five or more digits, are usually the city, state and zip code. By programming the application to look for a pattern, the information can be electronically identified, extracted, or otherwise utilized or manipulated.

*PCI:* Peripheral Component Interface (Interconnect). A high-speed interconnect local bus used to support multimedia devices.

*PCMCIA:* Personal Computer Memory Card International Association. Plug-in cards for computers (usually portables), which extend the storage and/or functionality.

*PDA (Personal Digital Assistant):* A small, usually hand-held, computer which “assists” business tasks.

*PDF (Portable Document Format):* An imaging file format technology developed by Adobe Systems. PDF captures formatting information from a variety of applications in such a way that they can be viewed and printed as they were intended in their original application by practically any computer, on multiple platforms, regardless of the specific application in which the original was created. PDF files may be text-searchable or image-only. Adobe® Reader, a free application distributed by Adobe Systems, is required to view a file in PDF format. Adobe® Acrobat, an application marketed by Adobe Systems, is required to edit, capture text, or otherwise manipulate a file in PDF format.

*Personal Computer (PC):* computer based on a microprocessor and designed to be used by one person at a time

*Petabyte (PB):* A unit consisting of 1,000 or 1,024 terabytes. *See* Byte.

*Phase Change:* A method of storing information on rewritable optical disc s.

*Physical Disc:* An actual piece of computer media, such as the hard disc or drive, floppy discs, CD-ROM discs, Zip discs, etc.

*Physical File Space:* When a file is created on a computer, a sufficient number of clusters (physical file space) are assigned to contain the file. If the file (logical file space) is not large enough to completely fill the assigned clusters (physical file space) then some unused space will exist within the physical file space. This unused space is referred to as file slack and can contain unused space, previously deleted/overwritten files or fragments thereof.

*PICA:* One sixth (1/6) of an inch. Used to measure graphics/fonts. There are 12 points per pica; 6 picas per inch; 72 points per inch.

*Picture Element:* The smallest addressable unit on a display screen. The higher the resolution (the more rows of columns), the more information can be displayed.

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

*PKI Digital Signature:* A document or file may be digitally signed using a party’s private signature key, creating a “digital signature” that is stored with the document. Anyone can validate the signature on the document using the public key from the digital certificate issued to the signer. Validating the digital signature confirms who signed it, and ensures that no alterations have been made to the document since it was signed. Similarly, an email message may be digitally signed using commonly available client software that implements an open standard for this purpose, such as Secure Multipurpose Internet Mail Extensions (S/MIME). Validating the signature on the email can help the recipient know with confidence who sent it, and that it was not altered during transmission. *See* Certificate.

*Plaintext:* The least formatted and therefore most portable form of text for computerized documents.

**Platter:** One of several components that make up a computer hard drive. Platters are thin, rapidly rotating discs 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 sub-divided into bytes.

**PMS (Pantone Matching System):** A color standard in printing.

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

**Pointer:** A pointer is an index entry in the directory of a disc (or other storage medium) that identifies the space on the disc in which an electronic document or piece of electronic data resides, thereby preventing that space from being overwritten by other data. In most cases, when an electronic document is “deleted,” the pointer is deleted, which allows the document to be overwritten, but the document is not actually erased.

**Portable Volumes:** A feature that facilitates the moving of large volumes of documents without requiring copying multiple files. Portable volumes enable individual CDs to be easily regrouped, detached and reattached to different databases for a broader information exchange.

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

**Preservation:** The process of ensuring retention and protection from destruction or deletion all potentially relevant evidence, including electronic metadata. *See also* Spoliation.

**Preservation Notice, Preservation Order:** *See* Legal Hold.

**Printout:** A printed version of text of data, another term for which is hard copy.

**Private Network:** A network that is connected to the Internet but is isolated from the Internet with security measures allowing use of the network only by persons within the private network.

**Privilege Data Set:** The universe of documents identified as responsive and/or relevant, but withheld from production on the grounds of attorney-client privilege or work product.

**Processing Data:** In the context of this document, synonymous with Image Processing.

**Production:** The process of delivering to another party, or making available for that party’s review, documents deemed responsive to a discovery request.

**Production Data Set:** The universe of documents identified as responsive to document requests and not withheld on the grounds of attorney-client privilege or work product.

**Production De-Duplication:** Removal of a document if multiple copies of that document reside within the same production set. For example, if two identical documents are both marked responsive, non-privileged, production de-duplication ensures that only one of those documents is produced. *See* De-Duplication.

**Production Number:** Often referred to as the BATES number. A sequential number assigned to every page of a production for tracking and reference purposes. Often used in conjunction with a suffix or

prefix to identify the producing party, the litigation, or other relevant information. *See also Bates Number.*

*Program:* *See Application and Software.*

*Properties:* Fields of electronic information, or “metadata,” associated with a record or document such as creation date, author, date modified, blind cc’s and date received.

*Protocol:* Defines a common series of rules, signals and conventions that allow different kinds of computers and applications to communicate over a network. One of the most common protocols for networks is called TCP/IP.

*Proximity Search:* For text searches, the ability to look for words or phrases within a prescribed distance of another word or phrase, such as “accident” within 5 words of “tire.”

*Public Network:* A network that is part of the public Internet.

*PST:* A Microsoft Outlook email store. Multiple .pst files may exist and contain archived email.

*QBIC (Query By Image Content):* An IBM search system for stored images which allows the user to sketch an image, and then search the image files to find those which most closely match. The user can specify color and texture – such as “sandy beaches” or “clouds”.

*Quality Control (QC):* Steps taken to ensure that results of a given task, product or service are of sufficiently high quality; the operational techniques and activities that are used to fulfill requirements for quality. In document handling and management processes, this includes image quality (resolution, skew, speckle, legibility, etc.), and data quality (correct information in appropriate fields, validated data for dates, addresses, names/issues lists, etc.).

*Quarter Inch Cartridge (QIC):* Digital recording tape, 2000 feet long, with an uncompressed capacity of 5 GB.

*Query:* A request for specific information from a database or other data set.

*Queue:* A sequence of items such as packets or print jobs waiting to be processed. For example, a print queue holds files that are waiting to be printed.

*RAID (Redundant Array of Independent Discs):* A method of storing data on servers that usually combines multiple hard drives into one logical unit thereby increasing capacity, reliability and backup capability. RAID systems may vary in levels of redundancy, with no redundancy being a single, non-mirrored disc as level 0, two discs that mirror each other as level 1, on up, with level 5 being one of the most common. RAID systems are more complicated to copy and restore.

*RAM (Random Access Memory):* Hardware inside a computer that retains memory on a short-term basis and stores information while the computer is in use. It is the “working memory” of the computer into which the operating system, startup applications and drivers are loaded when a computer is turned on, or where a program subsequently started up is loaded, and where thereafter, these applications are executed. RAM can be read or written in any section with one instruction sequence. It helps to have more of this “working space” installed when running advanced operating systems and applications. RAM content is erased each time a computer is turned off. (*See Dynamic Random Access Memory - DRAM*).

**Raster/Rasterized (Raster or Bitmap Drawing):** A method of representing an image with a grid (or “map”) of dots. Typical raster file formats are GIF, JPEG, TIFF, PCX, BMP, etc.

**Record:** Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business (ISO 15489(1)).

**Record Custodian:** A records custodian is an individual responsible for the physical storage and protection of records throughout their retention period. In the context of electronic records custodianship may not be a direct part of the records management function in all organizations. For example, some organizations may place this responsibility within their Information Technology Department, or they may assign responsibility for retaining and preserving records with individual employees.

**Record Lifecycle:** The time period from which a record is created until it is disposed.

**Record Owner:** The records owner is the subject matter expert on the contents of the record and is responsible for the lifecycle management of the record. This may be, but is not necessarily, the author of the record.

**Record Series:** A description of a particular set of records within a file plan. Each category has retention and disposition data associated with it, applied to all record folders and records within the category. (DOD 5015)

**Record Submitter:** The Record Submitter is the person who enters a record in an application or system. This may be, but is not necessarily, the author or the record owner.

**Record:** Information, regardless of medium or format that has value to an organization. Collectively the term is used to describe both documents and recorded data.

**Records Hold:** *See Legal Hold.*

**Records Management:** Records Management is the planning, controlling, directing, organizing, training, promoting, and other managerial activities involving the life-cycle of information, including creation, maintenance (use, storage, retrieval), and disposition, regardless of media.

**Records Manager:** The records manager is responsible for the implementation of a records management program in keeping with the policies and procedures that govern that program, including the identification, classification, handling and disposition of the organization’s records throughout their retention life. The physical storage and protection of records may be a component of this individual’s functions, but it may also be delegated to someone else. *See Records Custodian.*

**Records Retention Period, Retention Period:** The length of time a given records series must be kept, expressed as either a time period (*e.g.*, four years), an event or action (*e.g.*, audit), or a combination (*e.g.*, six months after audit).

**Records Retention Schedule:** A plan for the management of records listing types of records and how long they should be kept; the purpose is to provide continuing authority to dispose of or transfer records to historical archives.

**Records Store:** *See Repository for Electronic Records.*

**Recover, Recovery:** *See* Restore.

**Redaction:** A portion of an image or document is intentionally concealed to prevent disclosure of specific portions. Often done to avoid production of privileged or irrelevant materials.

**Refresh Rate:** The number of times per second a display (such as on a CRT or TV) is updated.

**Region (of an image):** An area of an image file that is selected for specialized processing. Also called a “zone.”

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

**Relative Path:** An implied path.

**Remote Access:** The ability to access and use digital information from a location off-site from where the information is physically located. For example, to use a computer, modem, and some remote access software to connect to a network from a distant location.

**Render Images:** To take a native format electronic file and convert it to an image that appears as the original format file as if printed to paper.

**Report:** Formatted output of a system providing specific information.

**Repository for Electronic Records:** Repository for Electronic Records is a direct access device on which the electronic records and associated metadata are stored. (DoD 5015) Sometimes called a “records store” or “records archive.”

**Residual Data:** Residual Data (sometimes referred to as “Ambient Data”) refers to data that is not active on a computer system. Residual data includes (1) data found on media free space; (2) data found in file slack space; and (3) data within files that has functionally been deleted in that it is not visible using the application with which the file was created, without use of undelete or special data recovery techniques. May contain copies of deleted files, Internet files and file fragments.

**Resolution:** *See* DPI.

**Restore:** To transfer data from a backup medium (such as tapes) to an on-line system, often for the purpose of recovery from a problem, failure, or disaster. Restoration of archival media is the transfer of data from an archival store to an on-line system for the purposes of processing (such as query, analysis, extraction, or disposition of that data). Archival restoration of systems may require not only data restoration but also replication of the original hardware and software operating environment. Restoration of systems is often called “recovery”.

**Retention Schedule:** *See* Records Retention Schedule.

**Reverse Engineering:** The process of analyzing a system to identify its intricacies and their interrelationships, and create depictions of the system in another form or at a higher level. Reverse engineering is usually undertaken in order to redesign the system for better maintainability or to produce a copy of a system without utilizing the design from which it was originally produced. For example, one might take the executable code of a computer program, run it to study how it behaved with different input, and then attempt to write a program which behaved the same or better.

*Review:* The culling process produces a dataset of potentially responsive documents which are then examined and evaluated for a final selection of relevant or responsive documents and assertion of privilege exception as appropriate. *Also see On-Line Review.*

*Rewriteable Technology:* Storage devices where the data may be written more than once – typically hard drives, floppies and optical discs.

*RFC822:* Standard that specifies a syntax for text messages that are sent among computer users, within the framework of email.

*RGB (Red, Green and Blue):* The three primary colors in the additive color family which create all the computer color video signals for a computer's color terminal.

*RIP:* The procedures used to unbundle email collections into individual emails during the e-discovery process while preserving authenticity and ownership.

*RIM:* Records and information management.

*RLE (Run Length Encoded):* Compressed image format; supports only 256 colors; most effective on images with large areas of black or white.

*ROM (Read Only Memory):* Random memory which can be read but not written or changed. Also, hardware, usually a chip, within a computer containing programming necessary for starting up the computer, and essential system programs that neither the user nor the computer can alter or erase. Information in the computer's ROM is permanently maintained even when the computer is turned off.

*Root Directory:* The top level in a hierarchical file system. For example on a PC, the root directory of your hard drive, usually C:, contains all the second-level subdirectories on that drive.

*Rotary Camera:* In microfilming, the papers are read “on the fly” with a camera that’s synchronized to the motion.

*Router:* A device that forwards data packets along networks. A router is connected to at least two networks, commonly two LANs or WANs or a LAN and its ISP's network. Routers are located at gateways, the places where two or more networks connect.

*Sampling Rate:* The frequency at which analog signals are converted to digital values during digitization. The higher the rate, the more accurate the process.

**Sampling:** Sampling usually (but not always) refers to the process of testing a database for the existence or frequency of relevant information. It can be a useful technique in addressing a number of issues relating to litigation, including decisions about what repositories of data are appropriate to search in a particular litigation, and determinations of the validity and effectiveness of searches or other data extraction procedures.

*SAN (Storage Area Network):* A high-speed subnetwork of shared storage devices. A storage device is a machine that contains nothing but a disc or discs for storing data. A SAN's architecture works in a way that makes all storage devices available to all servers on a LAN or WAN. As more storage devices are added to a SAN, they too will be accessible from any server in the larger network. In this case, the server merely acts as a pathway between the end user and the stored data. Because stored data does not reside

directly on any of a network's servers, server power is utilized for business applications, and network capacity is released to the end user. *Also see Network.*

**Scalability:** The capacity of a system to expand without requiring major reconfiguration or re-entry of data. For example, multiple servers or additional storage can be easily added.

**Scale-to-Gray:** An option to display a black and white image file in an enhanced mode, making it easier to view. A scale-to-gray display uses gray shading to fill in gaps or jumps (known as aliasing) that occur when displaying an image file on a computer screen. Also known as grayscale.

**Scanner:** An input device commonly used to convert paper documents into images. Scanner devices are also available to scan microfilm and microfiche.

**Scanning Software:** Software that enables a scanner to deliver industry standard formats for images in a collection. Enables the use of OCR and coding of the images.

**Schema:** A set of rules or conceptual model for data structure and content, such as a description of the data content and relationships in a database.

**Scroll Bar:** The bar on the side or bottom of a window that allows the user to scroll up and down through the window's contents. Scroll bars have scroll arrows at both ends, and a scroll box, all of which can be used to scroll around the window.

**SCSI (Small Computer System Interface):** Pronounced "skuzzy." A common, industry standard, electronic interface (highway) between computers and peripherals, such as hard discs, CD-ROM drives and scanners. SCSI allows for up to 7 devices to be attached in a chain via cables. As of this writing, the current SCSI standard is "SCSI II," also known as "Fast SCSI."

**SDLT (Super DLT):** A type of backup tape which can hold up to 220 GB or 330 CDs, depending on the data file format. *See DLT.*

**Search:** *See* Compliance Search, Concept Search, Contextual Search, Boolean Search, Full-Text Search, Fuzzy Search, Index, Keyword Search, Pattern Recognition, Proximity Search, QBIC, Sampling, *and* Search Engine.

**Search Engine:** A program that enables search for keywords or phrases, such as on web pages throughout the World Wide Web.

**Sector:** A sector is normally the smallest individually addressable unit of information stored on a hard drive platter, and usually holds 512 bytes of information. Sectors are numbered sequentially starting with 1 on each individual track. Thus, Track 0, Sector 1 and Track 5, Sector 1 refer to different sectors on the same hard drive. The first PC Hard discs typically held 17 sectors per track. Today, they can hold thousands of sectors per track.

**Serial Line Internet Protocol (SLIP):** A connection to the Internet in which the interface software runs in the local computer, rather than the Internet's.

**Serif:** The little cross bars or curls at the end of strokes on certain type fonts.

**Server:** Any central computer on a network that contains data or applications shared by multiple users of the network on their client PCs. A computer that provides information to client machines. For

example, there are web servers that send out web pages, mail servers that deliver email, list servers that administer mailing lists, FTP servers that hold FTP sites and deliver files to users who request them, and name servers that provide information about Internet host names. *See File Server.*

*Service-Level Agreement:* A service-level agreement is a contract that defines the technical support or business parameters that a service provider or outsourcing firm will provide its clients. The agreement typically spells out measures for performance and consequences for failure.

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

*SHA-1:* Secure Hash Algorithm, for computing a condensed representation of a message or a data file specified by FIPS PUB 180-1.

*Signature:* *See Certificate.*

*SIMM (Single, In-Line Memory Module):* A mechanical package (with “legs”) used to attach memory chips to printed circuit boards.

*Simplex:* One-sided page(s)

*Skewed:* Tilted images. *See De-skewing.*

*Slack/Slack Space:* The unused space on a cluster that exists when the logical file space is less than the physical file space. Also known as file slack. A form of residual data, the amount of on-disc file space from the end of the logical record information to the end of the physical disc record. Slack space can contain information soft-deleted from the record, information from prior records stored at the same physical location as current records, metadata fragments, and other information useful for forensic analysis of computer systems.

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

*SMTP (Simple Mail Transfer Protocol):* The protocol widely implemented on the Internet for exchanging email messages.

*Software application:* *See Application and Software.*

*Software:* Any set of coded instructions (programs) stored on computer-readable media that tells a computer what to do. Includes operating systems and software applications.

*Speckle:* Imperfections in an image as a result of scanning paper documents that do not appear on the original. *See De-speckling.*

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

*Spoliation:* Spoliation is the destruction of records which may be relevant to ongoing or anticipated litigation, government investigation or audit. Courts differ in their interpretation of the level of intent required before sanctions may be warranted. *See The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age, Guideline 3.*

*SPP (Standard Parallel Port):* *See Centronics.*

*SQL (Structured Query Language):* a standard fourth generation programming language (4GL - a programming language that is closer to natural language and easier to work with than a high-level language). The popular standard for running database searches (queries) and reports.

*Stand-Alone Computer:* A personal computer that is not connected to any other computer or network, except possibly through a modem.

*Standard Generalized Markup Language (SGML):* An informal industry standard for open systems document management which specifies the data encoding of a document's format and content.

*Status Bar:* A bar at the bottom of a window that is used to indicate the status of a task. For example, when an email message is sent, the status bar will fill with dots indicating that a message is being sent.

*Steganography:* The hiding of information within a more obvious kind of communication. Although not widely used, digital steganography involves the hiding of data inside a sound or image file. Steganalysis is the process of detecting steganography by looking at variances between bit patterns and unusually large file sizes.

*Storage Device:* A device capable of storing data. The term usually refers to mass storage devices, such as disc and tape drives.

*Storage Media:* See Magnetic or Optical Storage Media.

*Subjective Coding:* The coding of a document using legal interpretation as the data that fills a field, versus objective data that is readily apparent from the face of the document, such as date, type, author, addresses, recipients and names mentioned. Usually performed by paralegals or other trained legal personnel.

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

*Suspension Notice, Suspension Order:* See Legal Hold.

*SVGA (Super Video Graphics Adapter):* A graphics adapter one which exceeds the minimum VGA standard of 640 by 480 by 16 colors. Can reach 1600 by 1280 by 256 colors.

*Swap File:* A file used to temporarily store code and data for programs that are currently running. This information is left in the swap file after the programs are terminated, and may be retrieved using forensic techniques. Also referred to as a page file or paging file.

*System Administrator (sysadmin, sysop):* The person in charge of keeping a network working.

*System:* A system is: (1)a collection of people, machines, and methods organized to perform specific functions; (2)an integrated whole composed of diverse, interacting, specialized structures and sub-functions; and/or (3)a group of sub-systems united by some interaction or interdependence, performing many duties, but functioning as a single unit.

*T1:* A high speed, high bandwidth leased line connection to the Internet. T1 connections deliver information at 1.544 megabits per second.

*T3:* A high speed, high bandwidth leased line connection to the Internet. T3 connections deliver information at 44.746 megabits per second.

*Tape Drive:* A hardware device used to store or backup electronic data on a magnetic tape. Tape drives are usually used to back up large quantities of data due to their large capacity and cheap cost relative to other data storage options.

*Taxonomy:* The science of categorization, or classification, of things based on a predetermined system. In reference to Web sites and portals, a site's taxonomy is the way it organizes its data into categories and subcategories, sometimes displayed in a site map.

*TCP/IP (Transmission Control Protocol/Internet Protocol):* A collection of protocols that define the basic workings of the features of the Internet.

*Telephony:* Converting sounds into electronic signals for transmission.

*Templates, Document:* Sets of index fields for documents, providing framework for preparation.

*Temporary File - Temporary (or "temp"):* files are files stored on a computer for temporary use only, and are often created by Internet browsers. These temp files store information about Web sites that a user has visited, and allow for more rapid display of the Web page when the user revisits the site. Forensic techniques can be used to track the history of a computer's Internet usage through the examination of these temporary files. Temp files are also created by common office applications, such as word process or spreadsheet applications.

*Terabyte:* A unit of 1,000 or 1,024 gigabytes, or approximately a trillion bytes.

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

*Thin Client:* A networked user computer that acts only as a terminal and stores no applications or user files. May have little or no hard drive space. *See Client.*

*Thread:* A series of postings on a particular topic. Threads can be a series of bulletin board messages (for example, when someone posts a question and others reply with answers or additional queries on the same topic). A thread can also apply to chats, where multiple conversation threads may exist simultaneously.

*Thumb Drive:* *See Key Drive.*

*Thumbnail:* A miniature representation of a page or item for quick overviews to provide a general idea of the structure, content and appearance of a document. A thumbnail program may be standalone or part of a desktop publishing or graphics program. Thumbnails take considerable time to generate, but provide a convenient way to browse through multiple images before retrieving the one needed. Programs often allow clicking on the thumbnail to retrieve it.

*TIFF (Tagged Image File Format):* One of the most widely used and supported graphic file formats for storing bit-mapped images, with many different compression formats and resolutions. File name has .TIF extension. Can be black and white, gray-scaled, or color. Images are stored in tagged fields, and programs use the tags to accept or ignore fields, depending on the application. The format originated in the early 1980's.

*TIFF Group III (compression):*A one-dimensional compression format for storing black and white images that is utilized by many fax machines. *See* TIFF.

*TIFF Group IV (compression):*A two-dimensional compression format for storing black and white images. Typically compresses at a 20-to-1 ratio for standard business documents. *See* TIFF.

*Toggle:* A switch that is either on or off, and reverses to the opposite when selected.

*Tool Kit Without An Interesting Name (TWAIN):* A universal toolkit with standard hardware/software drivers for multi-media peripheral devices.

*Toolbar:* The row of buttons right below the menu that perform special functions quickly and easily.

*Topology:* The geometric arrangement of a computer system. Common topologies include a bus (network topology in which nodes are connected to a single cable with terminators at each end), star (local area network designed in the shape of a star, where all end points are connected to one central switching device, or hub), and ring (network topology in which nodes are connected in a closed loop; no terminators are required because there are no unconnected ends). Star networks are easier to manage than ring topology.

*Track:* Each of the series of concentric rings contained on a hard drive platter.

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

*Typeface:*There are over 10,000 typefaces available for computers. The general categories are: oldstyle (faces have slanted serifs, gradual thick to thin strokes and a slanted stress - the “O” appears slanted), modern (faces have thin, horizontal serifs, radical thick to thin strokes and a vertical street - the “O” does not appear to slant); slab serif (faces have thick, horizontal serifs, little or no thick-to-thin in the strokes and a vertical stress - the “O” appears vertical); sans serif (faces have no serifs), script (from elaborate handwriting styles to casual, freeform, unconnected letter forms), decorative unusual fonts (designed to be very different and attention getting).

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

*UMS:* Universal messaging system.

*Unallocated Space* - The area of computer media, such as a hard drive, that does not contain *normally accessible* data. Unallocated space is usually the result of a file being deleted. When a file is deleted, it is not actually erased, but is simply no longer accessible through normal means. The space that it occupied becomes unallocated space, i.e., space on the drive that can be reused to store new information. Until portions of the unallocated space are used for new data storage, in most instances, the old data remains and can be retrieved using forensic techniques.

*Unitization – Physical and Logical:* The assembly of individually scanned pages into documents. Physical Unitization utilizes actual objects such as staples, paper clips and folders to determine pages that belong together as documents for archival and retrieval purposes. Logical unitization is the process of human review of each individual page in an image collection using logical cues to determine pages that belong together as documents. Such cues can be consecutive page numbering, report titles, similar headers and footers and other logical indicators. This process should also capture document relationships, such as parent and child attachments. *See also* Attachment.

**UNIX:** A software operating system.

**Upgrade:** new or better version of some hardware or software.

**Upload:** To send a file from one computer to another via modem, network, or serial cable. With a modem-based communications link, the process generally involves the requesting computer instructing the remote computer to prepare to receive the file on its disc and wait for the transmission to begin.

**URI (Uniform Resource Indicators):** a URL is a URI.

**URL (Uniform Resource Locators):** the addressing system used in the World Wide Web and other Internet resources. The URL contains information about the method of access, the server to be accessed and the path of any file to be accessed. A URL looks like this: [http://thesedonaconference.org/publications\\_html](http://thesedonaconference.org/publications_html).

**URL:** See Address.

**User-Added Metadata:** Data or work product created by a user while reviewing a document, including annotations and subjective coding information.

**V.32bis:** The ITU (*see ITU*) standard for 14.4 kbs modem communications.

**V.34:** The proposed ITU (*see ITU*) standard for 28.8 kbs modem communications.

**Validate:** In the context of this document, confirm or ensure well grounded logic, and true and accurate determinations.

**VAR/VAD/VASD:** Value-Added Reseller/Value-Added Dealer/Value-Added Specialty Distributor. Companies or people who sell computer hardware or software *and* “add-value” in the process. Usually, the value added is specific technical or marketing knowledge and/or experience.

**VDT (Video Display Terminal):** generic name for all display terminals.

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

**Vendor-Added Metadata:** Data created and maintained by the electronic discovery vendor as a result of processing the document. While some vendor-added metadata has direct value to customers, much of it is used for process reporting, chain of custody and data accountability. Contrast with Customer-Added Metadata.

**Verbatim Coding:** Extracting data from documents in a collection in a way that matches exactly as the information appears in the documents.

**Version, Record Version:** A particular form or variation of an earlier or original record. For electronic records the variations may include changes to file format, metadata or content.

**Vertical De- Duplication:** A process through which duplicate data are eliminated within a single custodial or production data set. *See Content Comparison, File level Binary Comparison Horizontal De- duplication, Meta Data Comparison.*

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

**Video Electronics Standards Association (VESA):** concentrates on computer video standards.

**Video Scanner Interface:** A type of device used to connect scanners with computers. Scanners with this interface require a scanner control board designed by Kofax, Xionics or Dunord.

**Virus:** a self-replicating program that spreads by inserting copies of itself into other executable code or documents. A program into which a virus has inserted itself is said to be infected, and the infected file (or executable code that is not part of a file) is a host. Viruses are a kind of malware (malicious software). Viruses can be intentionally destructive, for example by destroying data, but many viruses are merely annoying. Some viruses have a delayed payload, sometimes referred to a bomb. The primary downside of viruses is uncontrolled self-reproduction, which desecrates or engulfs computer resources.

**Vital Record:** A record that is essential to the organization’s operation or to the reestablishment of the organization after a disaster.

**VoIP (Voice over Internet Protocol):** Telephonic capability across an IP connection; increasingly used in place of standard telephone systems.

**Volume:** A volume is a specific amount of storage space on computer storage media such as hard drives, floppy discs, CD-ROM discs, etc. In some instances, computer media may contain more than one volume, while in others, one volume may be contained on more than one disc.

**Volume Boot Sector -** When a partition is formatted to create a volume, a volume boot sector is created to store information about the volume. One volume contains the operating system and its volume boot sector contains code used to load the operating system when the computer is booted up.

**VPN (Virtual Private Network):** A secure network that is constructed by using public wires to connect nodes. For example, there are a number of systems that enable creation of networks using the Internet as the medium for transporting data. These systems use encryption and other security mechanisms to ensure that only authorized users can access the network and that the data cannot be intercepted.

**WAV:File extension name for Windows sound files .** “.WAV” files can reach 5 Megabytes for one minute of audio.

**Web Site:** A collection of Uniform Resource Indicators (URIs), including Uniform Resource Locators (URLs), in the control of one administrative entity. May include different types of URIs (*e.g.*, file transfer protocol sites, telnet sites, as well as World Wide Web sites). *See* URI and URL.

**Workflow, Ad Hoc:** A simple manual process by which documents can be moved around a multi-user review system on an “as-needed” basis.

**Workflow, Rule-Based:** A programmed series of automated steps that route documents to various users on a multi-user review system.

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

**WORM Discs:** Write Once Read Many Discs. A popular archival storage media during the 1980s. Acknowledged as the first optical discs, they are primarily used to store archives of data that cannot be altered. WORM discs are created by standalone PCs and cannot be used on the network, unlike CD-ROM discs.

**WORM (Write-Once, Read-Many):** Data storage devices (e.g. CD-ROM's) where the space on the discs can *only* be written *once*. The data is *permanently* stored. This is often today's primary media for archival information. Common disc sizes run from 5.25" (1.3 gigabytes) to 12" (8 to 10 gigabytes) capacities. There is also a 14" disc (13 to 15 gigabytes), only manufactured by Kodak's optical storage group. WORMs can also be configured into jukeboxes. There are various technologies. The expected viable lifetime of a WORM is at least 50 years. Since it's impossible to change, the government treats it just like paper or microfilm and it is accepted in litigation and other record-keeping applications. On the negative side, there is no current standard for how WORMs are written. The only ISO standard is for the 14" version, manufactured only by one vendor. A 5.25" standard is emerging from the European Computer Manufacturing Association but is not yet accepted. Further, WORM discs are written on both sides, but there are currently no drives that read both sides at the same time. As for speed, WORM is faster than tape or CD-ROM, but slower than magnetic. Typical disc access times run between 40 and 150 milliseconds (compared with 11 ms for fast magnetic discs and 300 ms for CD-ROM). Data transfer rates run between 1 and 2 MB/sec (compared with 5 to 10 for magnetic discs and 600KB/sec for CD-ROM).

**WWW (World Wide Web):** All of the computers on the Internet which use HTML-capable software (Netscape, Explorer, etc.) to exchange data. Data exchange on the WWW is characterized by easy-to-use graphical interfaces, hypertext links, images, and sound. Today the WWW has become synonymous with the Internet, although technically it is really just one component.

**WYSIWYG:** "What You See Is What You Get" – Display and software technology which shows on the computer screen exactly what will print. Often requires a large, high-density monitor.

**X.25:** A standard protocol for data communications.

**XML:** *See* Extensible Markup Language.

**Zip Drives:** A floppy disc drive that can hold a large amount of data, usually as much as 750 megabytes or more. Often used for backing up hard discs.

**ZIP:** A common file compression format that allows quick and easy storage for transport.

**Zone OCR:** An add-on feature of the imaging software that populates document templates by reading certain regions or zones of a document, and then placing the text into a document index.

- END -