wgs℠

THE SEDONA CONFERENCE® GLOSSARY:

E-Discovery & Digital Information Management (THIRD EDITION)

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

SEPTEMBER 2010 VERSION

The Sedona Conference®

THE SEDONA GLOSSARY:
for E-Discovery and Digital Information Management (Third Edition)

Editor:
Sherry B. Harris, Law Offices of Conor R. Crowley

Contributing Editor:
Paul H. McVoy, Milberg LLP

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

Editors and Contributors to the 2005 and 2007 Editions:[1]
Conor R. Crowley, Esq., Labaton Sucharow & Rudoff LLP; Sherry B. Harris, Hunton &
Williams LLP; Matthew I. Cohen, Esq., Skadden, Arps, Slate, Meagher & Flom; Megan E. Jones,
Esq., Cohen, Milstein, Hausfeld & Toll, PLLC; Anne Kershaw, Esq., A. Kershaw, PC//
Attorneys & Consultants; Mark V. Reichenbach, Merrill Lynch, OGC

[1]With affiliations at the time of their work on the 2007 Edition.

REPRINT REQUESTS:
Requests for reprints or reprint information should be directed to
Richard Braman, Executive Director of The Sedona Conference®,
at rgb@sedonaconference.org 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®.

wgs℠

Preface

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), American National Standards Institute (www.ansi.org), International Organization for Standardization (www.iso.org), U.S. National Archives & Records Administration (www.archives.gov), and other professional organizations.

The Sedona Conference® wishes to acknowledge the contributions of all the Working Group 1 and RFP+ Vendor Panel and User Group members who reviewed and commented on drafts of this Third Edition, as well as thank all of our Working Group Series^{SM} Sponsors, whose support is essential to the ability of The Sedona Conference® to develop Working Group Series^{SM} publications such as this. For a complete listing of our Sponsors, click on the "Sponsors" navigation bar on the home page of our Web site at www.thesedonaconference.org.

As with all of our publications, your comments are welcome. Please forward them to us at rgb@sedonaconference.org.

Richard G. Braman
Executive Director
The Sedona Conference®
Sedona, AZ
USA
September, 2010

The Sedona Conference® Glossary

Commonly Used Terms for E-Discovery and Digital Information Management

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 required to provide one or more witnesses to "testify as to matters known or reasonably available to the organization" on the topics requested by the notice. Sometimes the 30(b)(6) topics concern the discovery process itself, 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 disks, 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 method used by Lotus Notes developers to grant varying levels of access and user privileges within Lotus Notes databases.

ACM (Association for Computing Machinery): Professional association for computer professionals with a number of resources, including a special interest group on search and retrieval. See www.acm.org.

Active Data: Information residing on the direct access storage media (disk drives or servers) that is readily visible to the operating system and/or application software with which it was created. It is immediately accessible to users without restoration or reconstruction.

Active Records: Records related to current, ongoing, or in-process activities referred to on a regular basis to respond to day-to-day operational requirements. 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 a paper document.

Adware: See Spyware.

Agent: A program running on a computer that performs as instructed by a central control point to track file and operating system events and takes directed actions, such as transferring a file or deleting a local copy of a file, in response to such events.

AIIM: The Association for Information and Image Management, www.aiim.org. It focuses on ECM (enterprise content management).

Algorithm: A detailed formula or set of rules 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 non-control characters, such as @, #, $). Excludes control characters.

Ambient Data: See Latent Data and 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.

Annotation: 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, www.ansi.org - 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 that holds a 35mm frame of microfilm. Indexing information is punched in the card.

API (Application Programming Interface): Interface implemented by an application to enable interaction with another application. See MAPI.

Applet: a small program typically designed as an add-on to another program, allowing greater functionality for a specific purpose other than what the original program intended, e.g., a game applet for a Web browser.

Appliance: A prepackaged piece of hardware and software designed to perform a specific function on a computer network, for example, a firewall.

Application: 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.

Application Metadata: Data created by the application specific to the ESI being addressed, embedded in the file and moved with the file when copied; copying may alter application metadata. See also Metadata.

Application Service Provider (ASP): An Internet-based organization hosting software applications on its own servers within its own facilities. Customers license the application and access it over the Internet or via a private line connection. See SaaS.

Architecture: Refers to the hardware, software or combination of hardware and software comprising a computer system or network. "Open architecture" describes computer and network components that are more readily interconnected and interoperable. "Closed architecture" describes components that are less readily interconnected and interoperable.

Archival Data: 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: 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, both paper and electronic, 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 process with computers. It is also sometimes defined as solving 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," 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 to the width of any image. The aspect ratio of an image must be maintained to prevent distortion.

Attachment: A record or file associated with another record for the purpose of retention, transfer, processing, review, production, and routine records management. There may be multiple attachments associated with a single "parent" or "master" record. In many records and information management programs, or in a litigation context, the attachments and associated record(s) may be managed and processed as a single unit. In common use, this term often refers to a file (or files) associated with an email for retention and storage as a single Message Unit. See Document Family, Message Unit, and Unitization.

Attribute: A characteristic of data that sets it apart from other data, or property of a file aspect such as location, size, or type. The term attribute is sometimes used synonymously with "data element" or "property."

Audit Log or Audit Trail: 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 person, office, or designated position responsible for an item's 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, a software application producing a 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.

Avatar: A graphical representation of a user in a shared virtual reality, such as Web forums or chat rooms.

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 active 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 active ESI that 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 that impacts its suitability for certain purposes.

Backup Tape: Magnetic tape used to store copies of ESI, for use when restoration or recovery is required. Backup tapes typically use data compression, which increases restoration time and expense, given the lack of uniform standards governing data compression.

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.

Bandwidth: The amount of data a network connection can accommodate in a given period of time. Bandwidth is usually stated in kilobits per second (kbps), megabits per second (mps), or gigabytes per second (gps).

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

Batch File: A batch file is a set of one or more instructions that are created in a computer program to perform a particular type of computer system function (.BAT is the DOS batch file extension).

Batch Processing: The processing of a large amount of ESI in a single step.

Bates Number: Sequential numbering used to track documents and images in production data sets, where each page or file is assigned a unique production number. Often used in conjunction with a suffix or prefix to identify a 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.

Bayesian: Refers to the statistical approach of Thomas Bayes, an 18th century mathematician and clergyman. Bayes published a theorem that describes how to calculate conditional probabilities from the combinations of observed events and prior probabilities. Many information retrieval systems implicitly or explicitly use Bayes' probability rules to compute the likelihood that a document is relevant to a query.

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 (one of the largest and most active).

Beginning Document Number or BegDoc#: A unique number identifying the first page of a document or a number assigned to identify a native file.

Bibliographical/Objective Coding: Manually recording objective information from documents such as date, authors/recipients/carbon copies, blind copies, and associating the information with a specific document. See Coding and Indexing.

Binary: The Base 2 numbering system used in digital computing that 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, disk drives, and communication ports in addition to handling certain miscellaneous functions.

Bit: Binary digit - the smallest unit of computer data. A bit consists of either 0 or 1. There are eight bits in a byte. See Byte.

Bitmap: A Bitmap provides information on the placement and color of individual bits, as well as allows the creation of characters or images by creating a picture composed of individual bits (pixels).

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

Bitonal: A bitonal image uses only black and white.

BMP: A Windows file format for storing bitmap images.

BMS (BlackBerry® Messaging Service): A text message protocol for use between two or more BlackBerry® devices and addressed using the devices PIN, a unique address assigned to all BlackBerrys®. May also be referred to as BBM.

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

Boolean Search: Boolean Searches use logical operators such as "and", "or," and "not" to include or exclude terms from a search. See Natural Language Search.

Boot: To start up or reset a computer.

Boot Sector/Record: See Master Boot Sector/Record and Volumn Boot Sector/Record.

BPI (Bits Per Inch): A unit of measure of data densities in disk and magnetic tape systems.

bps (bits per second): A measurement of the rate of data transfer. See Bandwidth.

Broadband: Commonly used in the context of high bandwidth Internet access made available through a variety of quickly evolving technologies; the ability to divide bandwidth for sharing by multiple simultaneous signals increases transmission speed. Total available bandwidth may vary among providers.

Brontobyte: 1,024 yottabytes. See Byte.

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

Burn: The process of creating a copy of information onto a CD, DVD, or other storage media.

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 ("BPO"): 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, Exabyte, Zettabyte, Yottabyte, Brontobyte, and Geopbyte (listed here in order of increasing volume).

Cache: A dedicated, high speed storage location that can be used for the 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 storing of frequently-used data to speed access. See also Cache.

CAD (Computer Aided Design): The use of a wide range of computer-based tools that assist engineers, architects, and other design professionals in their design activities.

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. Also known as Cross Custodial De-Duplication, Global De-Duplication or Horizontal De-Duplication. See De-Duplication.

Catalog: See Index.

CCD (Charge Coupled Device): A computer chip of which the output 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: Consultative Committee for International Telephone & Telegraphy. Sets standards for phones, faxes, modems, etc. The standard exists primarily for fax 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.

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

CD-R, CD+R (Compact Disk Recordable): See Compact Disk.

CD-RW (Compact Disk Re-Writable): See Compact Disk.

CD-ROM (Compact Disk Read-Only Memory): See Compact Disk.

Certificate: An electronic affidavit vouching for the identity of the transmitter. See Digital Certificate, Digital Signature, PKI Digital Signature.

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

Chaff/winnowing: 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 or otherwise transferred or submitted; used to prove that evidence has not been altered or tampered with in any way; necessary both to assure admissibility and authenticity.

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

Checksum: A value used to ensure data is stored or transmitted without error. It is created by calculating the binary values in a block of data using some algorithm and storing the results with the data. When the data is retrieved from memory or received at the other end of a network, a new checksum is computed and matched against the existing checksum. A non-match indicates an error.

Child: See Document.

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

CIFS (Common Internet File System): Used for client/server communication within Microsoft® operating systems. With CIFS, users with different platforms and computers can share files without having to install new software.

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.

Clawback Agreement: An agreement outlining procedures to be followed to protect against waiver of privilege or work product protection due to inadvertent production of documents or data.

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. Also commonly used as synonymous with an email application, by reference to the Email Client.

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.

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

Cloud Computing: "[A] model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." http://csrc.nist.gov/groups/SNS/cloud-computing/ (last visited June 22, 2010). For further explanation see the NIST Web site cited.

Cluster (File): The smallest unit of storage space that can be allocated to store a file on operating systems. Windows and DOS organize hard disks based on Clusters (also known as allocation units), which consist of one or more contiguous sectors. Disks 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 bitmap: Used in NTFS (New Technology File System) to keep track of the status (free or used) of clusters on the hard drive. See NTFS.

Clustering: See Data Categorization.

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

Coding: Automated or human process by which documents are examined and evaluated using pre-determined codes, and the results recorded. 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. See Bibliographical/Objective Coding, Indexing, Level Coding, Subjective Coding, and Verbatim Coding.

COLD (Computer Output to Laser Disk): A computer programming process that outputs electronic records and printed reports to laser disk 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 and separate characters. Used in forms processing rather than boxes.

Comic Mode: Human-readable data, recorded on a strip of film that 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 Disk (CD): A type of optical disk storage media, compact disks come in a variety of formats. These formats include CD-ROM ("CD Read-Only Memory") - read-only; CD-R or CD+R ("CD Recordable") - can be written to once and are then read-only; and CD-RW ("CD Re-Writable") - can be written to multiple times.

Compliance Search: The identification of and search for 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.

Compound Document: A file that collects or combines more than one document into one, often from different applications, by embedding objects or linked data; multiple elements may be included, such as images, text, animation, or hypertext. See also OLE.

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 that do not preserve all of the original information, and are most commonly used for photographs, audio, and video. See Container File, Decompression, Lossless Compression, and Lossy Compression.

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

Computer Forensics: Computer Forensics 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).

Concatenate: Generally, to add by linking or joining to form a chain or series; two or more databases of similar structure can be concatenated to enable the user to reference them as one.

Concept search: The use of word meanings to identify documents relevant to a specific query. Word meanings can be derived from any of a number of sources, including dictionaries, thesauri, taxonomies, ontologies, or computed mathematically from the context in which the words occur. Concept searching typically improves the relevance ranking of the search results and can identify additional documents that are meaningfully related to the query even if they do not have the specific query term in them.

Conceptual Analytics: Using one or more of a number of mathematical algorithms or linguistic methodologies to group documents by their common themes or ideas..

Container File: A single file containing multiple documents and/or files, e.g., .pst, .nsf and .zip files. The file must be ripped or decompressed to determine volume, size, record count, etc., and to be processed for litigation review and production. See Decompression and Rip.

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

Contextual Search: Using one of a number of mathematical algorithms or linguistic methodologies to enlarge search results to include not only exact term matches but also matches where terms are considered in context of how and where they frequently occur in a specific document collection or more general taxonomy. For example, a search for the term "diamond" may bring back documents related to baseball but with no reference to the word diamond because they frequently occur within the same documents and therefore have a logical association.

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

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

Cookie: A text file containing tracking information such as dates and times of Web site visits, deposited by a Web site onto a user's computer. The text file is accessed each time the Web site is visited by a specific user and updated with browsing and other information. The main purpose of cookies is to identify users and possibly prepare customized Web site for them, including the personalization of advertising appearing on the Web sites.

Coordinated Universal Time (UTC): a high precision atomic time standard with uniform seconds defined by International Time and leap seconds announced at regular intervals to compensate for the Earth's slowing rotation and other discrepancies. Leap seconds allow UTC to closely track Universal Time, a time standard based not on the uniform passage of seconds, but on the Earth's angular rotation. Time zones around the world are expressed as positive or negative offsets from UTC. Local time is UTC plus the time zone offset for that location, plus an offset (typically +1) for daylight savings, if in effect. For example, 3:00 a.m. Mountain Standard Time = 10:00 UTC – 7. As the zero point reference, UTC is also referred to as Zulu time (Z). See also Normalization.

Corrupted File: A file damaged in some way, such as by a virus, by software or hardware failure, or degradation with the passage of time, so that it is partially or completely unreadable 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. Also known as Microprocessor.

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. See Checksum.

CRM (Customer Relationship Management) Application: Applications 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 Case De-Duplication and De-Duplication.

CRT (Cathode Ray Tube): The picture tube of older computer monitors or televisions, to be distinguished from newer "flat" LCD or plasma screens.

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: See Record Custodian and Record Owner.

Custodian De-Duplication: Culls a document to the extent multiple copies of that document reside within the same custodian's data set. Also known as Vertical De-duplication. 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 a hard drive that is located at the same head position. See Platter.

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

DAD (Digital Audio Disk): Another term for compact disk.

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, whether created automatically by the computer, such as log files, or created by a user, such as the information entered on a spreadsheet.

Data Categorization: The categorization and sorting of ESI - such as foldering by "concept," content, subject, taxonomy, etc. - through the use of technology - such as search and retrieval software or artificial intelligence - to facilitate review and analysis.

Data Collection: See Harvesting.

Data Controller (as used in the EU Data Protection Act): The natural or legal person who alone or jointly with others determines the purposes for which and the manner in which any Personal Data are to be processed.

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

Data Encryption Standard (DES): A form of private key encryption developed by IBM in the late 1970s.

Data Extraction: The process of parsing data from any electronic documents into separate fields such as Date Created, Date Last Accessed, and Text.

Data Field: See Field.

Data Filtering: The process of identifying for inclusion or exclusion of specific data based on specified parameters, such as date range, author(s), and/or keyword search terms.

Data Formats: The organization of information for display, storage, or printing. Data is sometimes 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 Mapping: A method for an organization to capture where its ESI is stored, physically and virtually, in what format it is stored, backup procedures in place, how the ESI moves and is used throughout the organization, information about accessibility of the ESI, retention and lifecycle management practices and policies, and identity of records custodians.

Data Mining: Data mining generally refers to knowledge discovery in databases (structured data); often techniques for extracting summaries and reports from databases and data sets. In the context of electronic discovery, this term often refers to the processes used to cull through a collection of ESI to extract evidence for production or presentation in an investigation or in litigation. In certain bar association ethics opinions, "mining metadata" or "mining an electronic document" has been used to refer to the act of searching for and viewing ("hidden") metadata embedded in an electronic document. See also Text Mining.

Data Processor (as used in the EU Data Protection Act): A natural or legal person (other than an employee of the Data Controller) who processes Personal Data on behalf of the Data Controller.

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

Data Subject (as used in the EU Data Protection Act): An individual who is the subject of Personal Data.

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, SHA1, or SHA2. See also Digital Fingerprint, File Level Binary Comparison, and Hash Coding.

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. 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.

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

Date/Time Normalization: See Normalization.

Date Created: The date a file was created on the media where it currently resides. If a file is moved to a new storage media, typically this date will reflect the date of the move.

Date Last Accessed: The date the file was lasted accessed, meaning last opened or moved, or even copied depending on the technology used to copy.

Date Last Modified: The date the file was last changed. The change is normally a substantive change to the document or its metadata and not a change to its storage location or an indication whether the file was opened and read.

Date Sent: The date on which an email was sent by the original author.

Date Received: The date on which an email was received.

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 Federal Rule of Evidence 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.

DDE (Dynamic Data Exchange): A form of interprocess communications used by Microsoft® Windows to support the exchange of commands and data between two simultaneously running applications.

DEB (Digital Evidence Bag): A standardized electronic "wrapper" or "container" for electronic evidence to preserve and transfer evidence in an encrypted or protected form that prevents deliberate or accidental alteration. The secure "wrapper" provides metadata concerning the collection process and context for the contained data.

Decompression: To expand or restore compressed data back to its original size and format. See Compression.

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

De-Duplication ("De-Duping"): The process of comparing electronic records based on their characteristics and removing or marking duplicate records within the data set. The methodology deployed and 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 Near 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: 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 overwritten or "wiped." Even after the data itself has 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 disk space that has been designated as available for reuse; the deleted file remains intact until it is overwritten.

Deletion: The process whereby data is removed from active files and other data storage structures on computers and rendered more 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.

De-NIST: The use of an automated filter program that screens files against the NIST list of computer file types to separate those generated by a system and those generated by a user. See NIST List.

Descenders: The portion of a character that 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 the working area of the display on an individual PC.

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

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

Digital: Information stored as a string of ones and zeros (numeric). Opposite of analog.

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

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 Signature: A way to ensure the identity of the sender, utilizing public key cryptography and working in conjunction with certificates. See Certificate, Digital Certificate, and PKI Digital Signature.

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."

Dirty Text: OCR output reflecting text as read by the OCR engine(s) with no clean-up.

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

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

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

Disk/Disk: Round, flat storage media with layers of material that 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 that may be potentially relevant to the issues at hand and/or that 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.

Diskwipe: 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: Information belonging to an organization that 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 disk 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 that can hold up to 80 GB depending on the data file format.

DMCA (Digital Millennium Copyright Act): United States copyright law to protect against copyright infringement of data, address rights, and obligations of owners of copyrighted material, and the rights and obligations of internet service providers on whose systems the infringing material may reside.

Document (or Document Family): A collection of pages or files produced manually or by a software application, constituting a logical single communication of information, but consisting of more than a single stand-alone record. Examples include a fax cover, the faxed letter, and an attachment to the letter - the fax cover being the "Parent," and the letter and attachment being a "Child." See also Attachment, Load File, Message Unit, and Unitization - Physical and Logical.

Document Date: For documents originating as paper, the document date is usually the date manually captured from the face of the document. The date is sometimes estimated based on the content of the document if no specific date can be found. Documents originating in electronic form have several dates associated them. See Also: Date Last Modified, Date Created, Date Last Accessed, Date Sent, and Date Received.

Document Imaging Programs: Software used to scan paper documents and to store, manage, retrieve, and distribute documents quickly and easily.

Document Metadata: Properties about the file stored in the file, 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 correspondence, memo, report, article, and others.

DoD 5015: Department of Defense standard addressing records management.

Domain: A sub-network of servers and computers. 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.

DOS: See MS-DOS.

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.

Double Byte Language: See Unicode.

Download: To move data from another location to one's own, usually over a network or the Internet; indicates that data is being transmitted from one location to another location. See Upload.

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 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 that is periodically "refreshed" or updated – as opposed to "static" RAM chips that 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 computer program that controls various devices such as the keyboard, mouse, monitor, etc., and integrates them with other programs on a computer.

Drop-Down Menu: A menu window that opens on-screen to display context-related options. Also called pop-up menu or pull-down menu.

DRM (Digital Rights Management): Technology that controls access to protected data.

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 Disk or Digital Versatile Disk): A plastic disk, like a CD, on which data can be written and read. DVDs can hold more information and can support more data formats than CDs. Formats include: DVD-R or DVD+R ("DVD Recordable") - written to once and are then read-only; and DVD-RW ("DVD Re-Writable") - can be written to multiple times.

ECM: Enterprise content management.

EDB: Standing for Exchange Database, it is a Microsoft® Exchange Server email container file.

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 ("E-Discovery"): The process of identifying, preserving, collecting, preparing, reviewing, and producing electronically stored information ("ESI") 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. Often referred to as ILM (information lifecycle management).

Electronic File Processing: See Processing Data.

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

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.

Electronically Stored Information: See ESI.

Electrostatic Printing: A process in which paper is exposed to an electron charge, causing toner to stick 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. See also En.

Email (Electronic Mail): An electronic means for sending, receiving, and managing communications via a multitude of different structured data applications (email client software), such as Outlook or Lotus Notes, or those often known as "webmail," such as Google Gmail or Yahoo Mail. See Email Message.

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. See Email Message.

Email Archiving: A systematic approach to recording, saving, and indexing copies of email messages in a structured format providing more centralized search and retrieval capabilities. Solutions can be on-premises or hosted, and design and capabilities vary widely. See Journaling.

Email Client: See Email (Electronic Mail).

Email Message: A document created or received via an electronic mail system. Any attachments that may be transmitted with the email message, such as word processing and other electronic documents, are not part of the email message, but are part of the Message Unit or Document Family.

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, e.g., blind copy addressees, received date. 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. See Metadata.

Email String: An electronic conversation between two or more parties via email. Also referred to as an email "thread." See Thread.

Email Store: Files containing message units. See Container Files, Message Unit, EDB, OST, PST, and NSF.

Embedded Metadata: Generally hidden, but an integral part of ESI, such as "track changes" or "comments" in a word processing file or "notes" in a presentation file. While some metadata is routinely extracted during processing and conversion for e-discovery, embedded data may not be. Therefore, it may only be available in the original, native file. See also Application Metadata and Metadata.

Embedded Object: An object embedded within another object, often appearing as an icon or hyperlink. See also Compound Document.

EML: Generic email format.

En: In any print, font or size is equal to the width of the letter "n" in that font and size. See also Em.

Encoding: To change or translate into code; to convert information into digital format. For software, encoding is used for video and audio references, like encoding analogue format into digital or raw digital data into compressed format.

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 that uniquely marks the end of a data file.

Ephemeral Data: Data that exists for a very brief, temporary period and is transitory in nature, such as data stored in RAM.

EPP (Enhanced Parallel Port): See Port.

EPS (Encapsulated PostScript): Uncompressed files for images, text, and objects. Can only be printed on printers with PostScript drivers.

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

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

ESI: As referenced in the United States Federal Rules of Civil Procedure, information that is stored electronically, regardless of the media or whether it is in the original format in which it was created, as opposed to stored in hard copy (i.e., on paper).

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

Evidentiary Image or Copy: See Forensic Copy.

Exabyte: –1,024 petabytes (approximately one billion gigabytes). See Byte.

Exchange Server: A server running Microsoft® Exchange messaging and collaboration software. It is widely used by enterprises using Microsoft® infrastructure solutions. Among other things, Microsoft® Exchange manages email, shared calendars, and tasks.

Expanded Data: See Decompression.

Export: Data extracted or taken out of one environment or application usually in a prescribed format, and usually used 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. See Disk Partition.

Extensible Markup Language (**XML**): 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 Negative: A result that is not correct because it fails to indicate a match where one exists.

False Positive: A result that is not correct because it indicates a match where there is none.

Fast Mode Parallel Port: See Port.

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.

FAX: Short for facsimile. A process of transmitting documents by scanning them to digital, converting to analog, transmitting over phone lines, 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 defined area of a storage medium used to record 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 or Field Delimiter: A character that separates the fields in an individual record. For example, the CSV format uses a comma as the field separator.

File: A collection of related data or information stored as a unit under a specified name on storage medium.

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 (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 Header: See Header.

File Level Binary Comparison: Method of de-duplication using the digital fingerprint (hash) of a file to compare the individual content and location of bytes in one file against those of another file. See also Data Verification, De-Duplication, Digital Fingerprint, and Hash coding.

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 configuration, 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 backup the network. See Server.

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

File Signature: See Digital Signature.

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

File System: The engine that an operating system or program uses to organize and keep track of ESI. More specifically, the logical structures and software routines used to control access to the storage on a hard disk 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 keeps track of what happens to data when the user deletes a file or subdirectory.

File System Metadata: Logs and other logistical information generated by an operating system to track the demographics (name, size, location, usage, etc.) of the ESI and, not embedded within, but stored externally from the ESI. See also Metadata.

File Table: See MFT.

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.

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 security programs and/or hardware that protect the resources of a private network from unauthorized access by users outside of an organization or user group. A firewall filters information to determine whether to forward the information toward its destination.

Filter (verb): See Data Filtering.

Flash Drive: See Key Drive.

Flash Memory: A type of computer memory used for storage of data to a physical disk. Commonly associated with small portable drives used in the transfer of data or with the memory cards used by digital cameras.

Flat File: A non-relational text-based file (i.e., a word processing document).

Flatbed Scanner: A flat-surface scanner that allows users to create a digital image of books and other hard copy documents or objects. See Scanner.

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

Folder: See Directory.

Forensic Copy: 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 or slack space on the media. Forensic copies are often called "images" or "imaged copies." See Bit Stream Backup and Mirror Image.

Forensics: The scientific examination and analysis of data held on, or retrieved from, ESI 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 specifications for the exchange of documents and/or data between parties during a legal dispute. Used to refer both to file format (e.g., native vs. imaged format with agreed-upon metadata and extracted text in a load file) and the media on which the documents are produced (paper vs. electronic). It should be noted that not all ESI may be conducive to production in either the native format or imaged format, and some other form of production may be necessary. Databases, for example, present such issues. See Load File and Native Format.

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): To make a drive ready for first use. Erroneously thought to "wipe" drive. Typically, only overwrites the File Allocation Table, but not the actual 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.

Fragmentation: 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 disks. 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 governs the transfer of files between computers over a network or the Internet. The terms FTP server or FTP site are commonly used to refer to a location to upload/download and exchange data, particularly in large volume.

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

Full Path: A file location description that includes the drive, starting or root directory, all attached subdirectories and ending with the file or object name. Often referred to as the Path Name.

Full-Text Indexing: The extraction and compilation of text from a collection of ESI. Text is gathered both from the body of the data and selected metadata fields. See Index.

Full-Text Search: The ability to search a full-text index of ESI for specific words, numbers and/or combinations or patterns thereof.

Fuzzy Search: The method of searching an index that allows for one or more characters in the original search terms to be replaced by wild card characters so that a more broad range of data hits will be returned. For example, a fuzzy search for "fell" could return "tell," "fall," or "felt".

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 (PAB).

Geopbyte: 1,024 brontobytes. See Byte.

Ghost: See Bit Stream Backup.

GIF (Graphics Interchange Format): A common native file format for storing images first originated by CompuServe, an Internet Service Provider, in 1987. Limited to 256 colors.

Gigabyte (GB): 1,024 megabytes. See Byte.

Global De-Deduplication: See Case De-Duplication.

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

GPS (Global Positioning System): A technology used to track the location of ground based objects using three or more orbiting satellites.

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

Grayscale: See Scale to Gray.

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 a computer system in order to steal, change, or destroy information.

Half Duplex: Transmission systems that can send and receive data between computers, but not at the same time.

Halftone: See Dithering.

Handshake: A transmission that occurs at the beginning of a communication session between computers to establish the technical format of the communication.

Hard Drive: The primary storage unit of a computer, consisting of one or more magnetic media platters on which digital data can be written and erased. See Platter.

Harvesting: The process of retrieving or collecting ESI from any media; an e-discovery vendor or specialist "harvests" ESI from computer hard drives, file servers, CDs, and backup tapes for processing and loading 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, 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 Disk 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.

Header: In information technology, descriptive data placed at the beginning of a file or section of data that in part identifies the file and some of its attributes. A header can consist of multiple fields, each containing its own value. In email it is the part of the message containing information about the message, such as the sender, date sent and other brief details.

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 readily visible in the file directory. 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 ESI duplicated across multiple custodians or other production data sets normally by employing hash algorithms. See Case De-Duplication and De-Duplication.

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

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

HRS: Handwriting recognition software for interpreting handwriting into machine readable form.

HTCIA (High Technology Crime Investigation Association): Computer forensics non-profit association; resources include educational programs and list services.

HTML: HyperText Markup Language, developed by CERN of Geneva, Switzerland. The document format used on the Internet. (HTML+ adds support for multimedia.) 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. HTML is a markup or "presentation" language, not a programming language. Programming code can be imbedded in an HTML page to make it interactive. See Java.

HTTP(S) (HyperText Transfer Protocol): The underlying protocol used by the World Wide Web. HTTP defines how messages are formatted and transmitted, and what actions 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. HTTPS adds a layer of encryption to the protocol to protect the information that is being transmitted and is often used by Application Service Providers to protect the data being viewed over the Web.

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

Hyperlink: A link - usually appearing as an underlined or 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.

IaaS: (Infrastructure as a Service): a form of Cloud Computing that offers users computer infrastructure remotely. Specific services may include servers, software, or network equipment resources that can be rented on an as-needed basis without the cost of the devices or the resources needed to support them. See Cloud Computing.

Icon: In a GUI, a picture or drawing that 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 that define the rules of and algorithms for conversion, helpful for interpreting handwritten text. See HRS and OCR.

IDE (Integrated Drive Electronics): An engineering standard for interfacing computers and hard disks.

IEEE (Institute of Electrical and Electronic Engineers): An international association that advocates the advancement of technology as it relates to electricity. IEEE sponsors meetings, publishes a number of journals, and establishes standards.

ILM: Information lifecycle management.

Image: (1) 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. See Bit Stream Backup, Forensic Copy and Mirror Image. (2) An electronic or digital picture of a document (e.g., TIFF, PDF, etc.). See Image Processing, Processing Data, and Render Images.

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 a 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 convert data from its current native format to a fixed image for the purposes of preserving the format of a document and facilitating the transfer between parties, typically with the addition of a Bates Number to the face of each image. See Form of Production, Native Format, Processing Data, and Render Images.

Import: Data brought into an environment or application that 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. See Coding.

Index: A searchable catalog of information to maximize storage efficiency and allow for improved search. Also called "catalog." See Full Text Indexing.

Indexing: (1) The process of organizing data in a database to maximize storage efficiency and optimize searching; (2) Objective coding of documents to create a list similar to a table of contents. See Coding.

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. Instant Messages differ from email in their limited metadata and in that messages are often transitory and not stored past the messaging session.

Interlaced: To refresh a display every other line once per 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 that are completely refreshed less than 30 times per second. TV & CRT pictures must constantly be "refreshed."

Interleave: To arrange data in a noncontiguous way to increase performance. When used to describe disk drives, it refers to the way sectors on a disk 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 disk drive more efficient. The disk drive can access only one sector at a time, and the disk 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 www.itu.int.

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 (FTP), and Internet Relay Chat (IRC). Also known as "the net," "the information superhighway," and "cyberspace."

Internet Publishing Software: Specialized software that allows materials to be published to the Internet. The term Internet Publishing is sometimes used to refer to the industry of online digital publication as a whole.

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. See Track and Partition.

Intranet: A secure private network that uses Internet-related technologies to provide services within an organization or defined infrastructure.

IP address (Internet Protocol address): A unique name that identifies the physical location of a server on a network, expressed by a numerical value. See TCP/IP (e.g., 128.24.62.1).

IPX/SPX: Communications protocol used by Novell networks.

IRC (Internet Relay Chat): System allowing internet users to chat in real time.

IS/IT Information Systems or Information Technology: Usually refers to the department of an entity which designs, maintains, and assists users with regard to the computer infrastructure.

ISA: Industry Standard Architecture.

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

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

ISO (International Organization for Standardization): A worldwide federation of national standards bodies; www.iso.org.

ISO 8859-1: Also called Latin-1. A standard character encoding of the Latin alphabet used for most Western European languages. ISO 8859-1 is considered a legacy encoding in relation to Unicode, yet it is nonetheless still in common use today. The ISO 8859-1 standard consists of 191 printable characters from the Latin script. It is essentially a superset of the ASCII character encoding and a subset of the Windows-1252 character encoding. See ASCII and Windows-1252.

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

ISO 15489-1: The ISO standard addressing standardization of international best practices in records management.

ISP (Internet Service Provider): A business that provides access to the Internet, usually for a monthly fee.

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 that runs at scheduled intervals to manage business information by deleting, transferring, or archiving on-line data (such as email) that 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 Web sites.

Jaz (or Jazz) Drive: A removable disk drive that uses compression technology to store large amounts of data. A Jaz drive holds up to 2 GB of data. The drives are no longer manufacured.

JMS: Jukebox Management Software. See Jukebox.

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 electronic communication 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 email archiving solutions, that use proprietary storage formats designed to reduce the amount of storage space required. Journaling systems may also lack the sophisticated search and retrieval capabilities available with many email archiving solutions. See Email Archiving.

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 disks and loads them into a drive.

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. Key drives are also known as keychain drive, thumb drive, jump drive, and/or USB flash drive. Can be imaged and may contain residual data. Metadata detail may not be the equivalent of ESI maintained in more robust storage media.

Key Field: Database field(s) used for document searches and retrieval, and which are unique to the individual database record.

Keyword: Any specified word, or combination of words, used in a search, with the intent of locating certain results.

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, coaxial cable, or wireless transmission. See Network.

Landscape Mode: A page orientation or display such that the width exceeds the height (Horizontal).

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

Laser Printing: A printing process by which 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 disk (or jukebox), including the time to physically position the media under the read/write head, seek the correct address, and transfer it.

Latent Data: Deleted files and other ESI 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. Also known as ambient data. See Residual Data.

Latent Semantic Indexing and Analysis: A method of processing data that identifies relationships between data sets by analyzing terms and term frequency. Common applications include grouping documents together based on the documents' concepts and meanings instead of by simple searching.

Latin-1: See ISO 8859-1.

LCD (Liquid Crystal Display): Two polarizing transparent panels with a liquid crystal surface between; application of voltage to certain areas causes the crystal to turn dark, and a light source behind the panel transmits though crystals not darkened.

Leading: The amount of space between lines of printed text.

Legacy Data, Legacy System: Legacy Data is ESI in which an organization may have invested significant resources, 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 reasonably anticipated litigation, audit, government investigation, or other such matter that suspends the normal disposition or processing of records. Legal holds may encompass procedures affecting data that is accessible as well as data that is not reasonably accessible. 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." See The Sedona Conference® Commentary on Legal Holds, August 2007 Public Comment Version, available for download at www.thesedonaconference.org.

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. See Coding.

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. 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 or electronically processed files, and indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends. A load file may also contain data relevant to the individual documents, such as selected metadata, coded data, and extracted text. Load files should be obtained and provided in prearranged or standardized formats to ensure transfer of accurate and usable images and data.

Local Area Network (LAN): See LAN.

Locale: A set of parameters that define language, country, and any special system configurations that correspond to the language and country. For example, locale typically determines the date format (month first in the U.S., day first in the U.K.), the time format (12-hour clock in the US, 24-hour clock in some European countries), the keyboard layout, and so forth. These settings can be overridden, but the locale sets the default.

Log File: Created by an electronic device or application to record activity of a server, Web site, computer, or other domain.

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 Unitization: See Unitization - Physical and Logical.

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. A lossy image is one where the image after compression is different from the original image due to lost information. The differences may or may not be noticeable, but a lossy conversion process does not retain all the original information. JPEG is an example of a lossy compression method.

Lotus Domino: An IBM server product providing enterprise-level email, collaboration capabilities, and custom application platform; began life as Lotus Notes Server, the server component of Lotus Development Corporation's client-server messaging technology. Can be used as an application server for Lotus Notes applications and/or as a Web server. Has a built-in database system in the format of .NSF.

Lotus Notes: See Lotus Domino.

LPI (lines per inch): The number of lines in an inch, as found on screens that create halftones and four-color process images. The more lines per inch, the more detailed the image. With the growth of computer-generated imagery, the term DPI is quickly replacing the term LPI.

Lumen: Measure of brightness often associated with the amount of light output of a projector.

LTO (Linear Tape-Open): A type of magnetic backup tape that can hold as much as 800 GB of data, or 1200 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 stored. In email systems, each user has a separate mailbox. When the server receives email, the mail system automatically puts it in the appropriate mailbox.

Make-Available Production: Process by which a generally large universe of potentially responsive documents is made available to a requestor; the requestor selects or tags desired documents, and the producing party produces only the selected documents. See also Quick Peek.

Malware: Any type of malicious software program, typically installed illicitly, including viruses, Trojans, worms, key loggers, spyware, adware and others.

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

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

Marginalia: Handwritten notes on documents.

Master Boot Sector/Record: The 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.

Mastering: Making many copies of a disk from a single master disk.

MBOX: The format in which email is stored on traditional UNIX email systems.

MCA (Micro Channel Architecture): IBM bus standard rendered obsolete by the PCI bus.

MDE (Magnetic Disk 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 5) A hash algorithm used to give a numeric value to a digital file or piece of data. Commonly used in electronic discovery to find duplicates in a data collection. See Hash.

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

Megabyte (MB): 1,024 kilobytes. 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, disks, CDs, DVDs, key drives, and hard drives. 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: The text portion of an email that contains routing information of the email and may include author, recipient, and server information, which tracks the path of the email from its origin server to its destination mailbox.

Message Unit: An email and any attachments associated with it.

Metadata: Data typically stored electronically that describes characteristics of ESI, found in different places in different forms. Can be supplied by applications, users or the file system. Metadata can describe how, when, and by whom ESI was collected, created, accessed, modified, and how it is formatted. Can be altered intentionally or inadvertently. Certain metadata can be extracted when native files are processed for litigation. 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 to paper or electronic image. See also Application Metadata, Document Metadata, Email Metadata, Embedded Metadata, File System Metadata, User-Added Metadata, and Vendor-Added Metadata. For a more thorough discussion, see The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age (Second Edition).

Metadata Comparison: A comparison of specified metadata as the basis for de-duplication without regard to content. See De-Duplication.

MFT (Master File Table): The primary record of file storage locations on a Microsoft® Windows-based computer employing NTFS filing systems.

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

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

Microprocessor: See CPU.

Microsoft® Outlook: A personal information manager from Microsoft®, part of the Microsoft® Office suite. Although often used mainly as an email application, it also provides calendar, task and contact management, note taking, a journal, and Web browsing. Can be used as a stand-alone application, or operate in conjunction with Microsoft® Exchange Server to provide enhanced functions for multiple users in an organization, such as shared mailboxes and calendars, public folders, and meeting time allocation.

Microsoft® Outlook Express: A scaled down version of Microsoft® Outlook.

MiFi: A personal wireless hub that allows other users with the correct credentials to access the internet if within range of the hub.

Migrated Data: ESI that has been moved from one database or format to another.

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

Mirror Image: A bit by bit copy of any storage media. Often used to copy the configuration of one computer to anther computer or when creating a preservation copy. See Forensic Copy and Image.

Mirroring: The duplication of ESI for purposes of backup or to distribute Internet or network traffic among several servers with identical ESI. See also Bit Stream Backup, Disk Mirroring, and Image.

MIS: Management Information Systems.

MMS (Multimedia Messaging Service): A form of message that includes multimedia content such as pictures, video, or sounds used to communicate over mobile networks.

Modem: (Modulator-Demodulator) A device that can encode digital information into an analog signal (modulates) or decode the received analog signal to extract the digital information (demodulate).

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 ESI 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 ESI from mounted media to another media or to an on-line system.

MPEG-1, -2, -3, and -4: 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. MPEG-3 refers to the playing of CD clips.

MS-DOS: Microsoft®-Disk Operating System. Used in Windows-based personal computers as the control system prior to the introduction of 32-bit operating systems.

MSG: Generic format in which emails can be saved.

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 that 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 may be converted to a neutral format as part of the record acquisition or archive process. "Static" format (often called "imaged format"), 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 unless agreed-upon metadata and extracted text are preserved. In the conversion to static format, the metadata can be processed, preserved, and electronically associated with the static format file. However, with technology advancements, tools and applications are increasingly available to allow viewing and searching of documents in their native format, while still preserving pertinent metadata. It should be noted that not all ESI may be conducive to production in either the Native Format or imaged format, and some other form of production may be necessary. Databases, for example, often present such issues. See Form of Production and Load File.

Native Format Review: Review of ESI in its current "native" format using either an application capable of supporting native format review or the original application in which the ESI was created.

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"). See Boolean Search.

Near Duplicates: (1) A term used to describe two or more files that are similar to a certain percentage, for example files that are 90% similar are identified as Near Duplicates; used for review to locate similar documents and review at one time; (2) A term used to describe the longest email in an email conversation where the subparts are identified and suppressed in an email collection to reduce review volume.

Near-Line Data Storage: A term used to refer to a data storage system where data is not actively available to users, but is available through an automated system that enables the robotic retrieval of removable storage media or tapes. 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: A group of two or more computers and other devices connected together ("networked") for the exchange and sharing of 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 PCs or other devices, remote to each other, connected by telecommunications lines. Typically, a WAN may connect two or more LANs together.

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

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.

NIST List: A hash database of computer files developed by NIST to identify files that are system generated and generally accepted to have no substantive value in most instances.

NOC (Network Operations Center): The location where a network or computer array is monitored and maintained.

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

Non-Apparent Data: Data not normally seen on a printed version of ESI - whether "printed" to paper or image, such as tiff or pdf, e.g., spreadsheet formulas. See Embedded Metadata and Metadata.

Non-Interlace: When each line of a video image is scanned separately. Older CRT computer monitors use non-interlaced video.

NOS (Network Operating System): See Operating System.

Normalization: The process of reformatting data so that it is stored in a standardized form, such as setting the date and time stamp of a specific volume of ESI to a specific zone, often GMT, to permit advanced processing of the ESI, such as de-duplication. See also Coordinated Universal Time.

Notes Server: See Lotus Domino.

NSF: Lotus Notes container file (i.e., database.nsf); can be either an email database or the traditional type of fielded database. See Lotus Domino.

NTFS (New Technology File System): A high-performance and self-healing file system proprietary to Microsoft®, used in Windows NT, Windows 2000, Windows XP, and Windows Vista Operating Systems, that supports file-level security, compression and auditing. It also supports large volumes and powerful storage solution such as Redundant Array of Inexpensive Disks (RAID). An important feature of NTFS is the ability to encrypt files and folders to protect sensitive data.

Object: In personal computing, an object is a representation of something that a user can work with to perform a task and can appear as text or an icon. In a high-level method of programming called object-oriented programming (OOP), an object is a freestanding block of code that defines the properties of some thing.

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. See HRS and ICR.

Official Record Owner: See Record Owner.

Off-Line Data: The storage of ESI 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: ESI maintained or archived on removable disk (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 disk 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 that allows the linking of different files, or parts of files, together into one file without forfeiting any of the original file's attributes or functionality. See also Compound Document.

On-Line Review: The review of data on a computer, either locally on a network or via the Internet.

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

Online/On-Line: Connected to a network or the Internet.

Ontology: A collection of categories and their relationships to other categories and to words. An ontology is one of the methods used to find related documents when given a specific query.

Operating System (OS): An Operating System provides the software platform that directs the overall activity of a computer, network or system, and on which all other software programs and applications 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 disk and controlling peripheral devices such as disk 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 Disks: Computer media similar to a compact disk that cannot be rewritten. An optical drive uses a laser to read the ESI.

Optical Jukebox: See "Jukebox."

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

Outlook: See Microsoft® Outlook.

Over-inclusive: When referring to data sets returned by some method of query, search, filter, or cull, results that are returned overly broad.

Overwrite: To record or copy new data over existing data, as in when a file or directory is updated.

Paas: (Platform as a Service) A form of Cloud Computing which describes the outsourcing of the computer platform upon which development and other workflows can be performed without the costs of hardware, software and personnel. See Cloud Computing.

PAB (Personal Address Book): A Microsoft® Outlook list of contacts created and maintained by an individual user for personal use.

PackBits: A compression scheme that originated with the Macintosh®. Suitable only for black and white.

Packet: A unit of data sent across a network that may contain identity 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 method to temporarily store data outside of the main memory, but quickly retrievable. 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.

Parallel Port: See Port.

Parent: See Document.

Parsing: Transforms input text into a data structure suitable for later processing, while capturing the implied hierarchy of the input. Data may be parsed from one source of ESI to another.

Partition: 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: Indicates each logical volume contained on a disk 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 disk 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: (1) The hierarchical description of where a directory, folder, or file is located on a computer or network; (2) 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 ESI 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 Interconnect (Interface). 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) that extend the storage and/or functionality.

PDA (Personal Digital Assistant): A small, usually hand-held, computer used to perform communication and organizational tasks remotely.

PDF (Portable Document Format): A file format technology developed by Adobe Systems to facilitate the exchange of documents between platforms regardless of originating application by preserving the format and content.

Peer to Peer or P2P: A form of network organization that uses portions of each users resources, like storage space or processing power, for use by others on the network. Notorious examples include the storage sharing of Napster or Bittorrent.

Peripheral: Any accessory device attached to a computer, such as a disk drive, printer, modem, or joystick.

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

Personal Data (as used in the EU Data Protection Act): Data which relate to a natural person who can be identified from those Data, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural, or social identity. Also referred to as PII (Personally Identifiable Information).

Petabyte (PB): –1,024 terabytes (approximately one million gigabytes). See Byte.

PFC (Personal Filing Cabinet): The AOL proprietary email storage container file used for the local storage of emails, contacts, calendar events, and other personal information.

Phase Change: A method of storing information on rewritable optical disks.

Physical Disk: An actual piece of computer media, such as the hard disk or drive, floppy disks, CD-ROM disks, Zip disks, 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.

Physical Unitization: See Unitization - Physical and Logical.

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.

PII (Personally Identifiable Information): See Personal Data.

Ping: Executable command, used as a test for checking network connectivity.

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

PKI (Public Key Infrastructure) Digital Signature: A system, including hardware, software and policies, designed to manage digital certificates and match those certificates to specific users so that data can be validated as authentic. See Certificate, Digital Certificate, and Digital Signature.

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

Plasma Display: A type of flat panel display commonly use for large televisions in which many tiny cells are located between two panels of glass holding an inert mixture of gases that are then electronically charged to produce light.

Platter: One of several components that make up a computer hard drive. Platters are thin, rapidly rotating disks 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.

Plug and Play ("PNP"): A method by which new hardware may be detected, configured, and used by existing systems upon connection with little or no user intervention.

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.

Pointer: An index entry in the directory of a disk (or other storage medium) that identifies the space on the disk 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, that allows the document to be overwritten, but the document is not actually erased.

Port: An interface between a computer and other computers or devices, that can be divided into two primary groups based on signal transfer: serial ports send and receive one bit at a time via a single wire pair, while parallel ports send multiple bits at the same time over several sets of wires. See also USB Port. Software ports are virtual data connections used by programs to exchange data directly instead of going through a file or other temporary storage locations; the most common types are TCP and UDP.

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 page orientation or display such that the height exceeds the width (vertical).

Preservation: The process of retaining documents and ESI, including document metadata, for legal purposes and should include suspension of normal document destruction policies and procedures. See also Spoliation.

Preservation Notice, Preservation Order: See Legal Hold.

Printout: Printed data, also known as 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 legal privilege, a log of which is usually required to notify of withheld documents and the grounds on which they were withheld (e.g., work product, attorney-client privilege).

Process/Processing (as used in the EU Data Protection Act): Any operation or set of operations which is performed upon Personal Data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure, or destruction.

Processing Data: An automated computer workflow where native data is ingested by any number of software programs designed to extract text and selected metadata and then normalize the data for packaging into a format for the eventual loading into a review platform. May also entail identification of duplicates/de-duplication and rendering of data into delimited format.

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

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

Production Number: Often referred to as the "Bates" Number. A sequential number assigned to every page of a production for fixed image production formats, or to every file in a native file production, used 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: File level metadata describing attributes of the physical file, i.e., size, creation data, and author. See Metadata.

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.

Protodigital: Primitive or first-generation digital. Applied as an adjective to systems, software, "documents," or ways of thinking. The term was first used in music to refer to early computer synthesizers that attempted to mimic the sound of traditional musical instruments, and to early jazz compositions written on computers with that instrumentation in mind. In electronic discovery, this term is most often applied to systems or ways of thinking that — on the surface — appear to embrace digital technology, but attempt to equate ESI to paper records, ignoring the unique attributes of ESI. When someone says, "What's the big deal with e-discovery? Sure we have a lot of email. You just print it all out and produce it like you used to," that is an example of protodigital thinking. When someone says, "We embrace electronic discovery. We scan everything to .PDF before we produce it," that person is engaged in protodigital thinking — attempting to fit ESI into the paper discovery pardigm.

Proximity Search: A search performed to find two or more words within a specified distance from each other

PST: A Microsoft® Outlook email storage file containing archived email messages in a compressed format.

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

QBIC (Query By Image Content): An IBM search system for stored images that 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."

QR Code: A small, square matrix pattern of dots that can be read by an optical scanner or mobile phone camera. QR codes can store thousands of alphanumeric characters and may be affixed to business cards, advertising, product parts, or other objects in order to convey an address, an internet URL, or text information.

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: An electronic search request for specific information from a database or other ESI.

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.

Quick Peek: An initial production whereby documents and/or ESI are made available to the opposing party before being reviewed for privilege, confidentiality, or privacy, requiring stringent guidelines and restrictions to prevent waiver. This can be done under an agreement protecting against privilege and confidentiality waiver during the initial quick peek; 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. See also Make Available.

RAID (Redundant Array of Independent Disks): 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 disk as level 0, two disks that mirror each other as level 1, on up, with level 5 being one of the most common. RAID systems are more complicated to restore and copy.

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., and typically have jagged edges.

Record: (1) Information, regardless of medium or format that has value to an organization. (2) A single row of information or subset of data elements in a database.

Record Custodian: 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. See Record Owner.

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

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

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. See DOD 5015.

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

Records Archive: See Repository for Electronic Records.

Records Hold: See Legal Hold.

Records Management: 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 person 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-cycle. 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 Record Custodian.

Records Retention Period, Retention Period: The length of time a given record 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. Usually accomplished by applying an overlay. Often done to protect privileged or irrelevant portions of the document, including highly confidential, sensitive, or proprietary information.

Refresh Rate: The number of times per second a computer 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: (1) In document coding, it is the process of lining up an image of a form to determine the location of specific fields. See Coding. (2) Entering pages into a scanner such that they are correctly read.

Relative Path: The electronic path on a network or computer to an individual file from a common point on the network.

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 if the original format file were printed to paper. See Image Processing.

Report: Formatted output of a system providing specific information.

Repository for Electronic Records: A direct access device on which the electronic records and associated metadata are stored. Sometimes called a "records store" or "records archive."

Residual Data: Sometimes also referred to as "Ambient Data"; data that is not active on a computer system as the result of being deleted or moved to another location. 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. See Latent Data.

Resolution: Refers to the sharpness and clarity of an image. The term is most often used to describe monitors, printers, and graphic images. See DPI.

Restore: To transfer data from a backup medium (such as tapes) to an active 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 active 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 that behaved the same or better.

Review: The process of reading or otherwise analyzing documents to make a determination as to the document's applicability to some objective or subjective standard. Often used to describe the examination of documents in a legal context for their responsiveness to specific issues in a matter. See also On-Line Review.

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

RFC822: Standard that specifies a syntax for text messages that are sent between one or more 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: To extract ESI from container files, for example to unbundle email collections into individual emails, during the e-discovery process while preserving metadata, authenticity, and ownership. Also used to describe the extraction or copying of data to or from an external storage device.

RIM (Records and Information Management): Also the name of the company that developed and sells Black Berry® devices; short for Research In Motion

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 that 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 the hard drive, usually C:, contains all the second-level subdirectories on that drive.

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 ISPs network. Routers are located at gateways, the places where two or more networks connect. See also Wireless Router.

RTF (Rich Text Format): A standard text file format that preserves minimal stylistic formatting of document files for ease in exchange between various parties with different software.

SaaS (Software as a Service): Software application delivery model where a software vendor develops a Web-native software application and hosts and operates (either independently or through a third-party) the application for use by its customers over the Internet. Customers pay not for owning the software itself but for using it. See Application Service Provider and Cloud Computing.

Sampling: Sampling usually refers to the process of testing a database or a large volume of ESI 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 sub-network of shared storage devices. A storage device is a machine that contains nothing but a disk or disks 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. 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. See also Network.

SAS-70 (Statement on Auditing Standards No. 70, Service Organizations): An auditing standard developed by the American Institute of Certified Public Accountants (AICPA), which includes an examination of an entity's "controls" over information technology, security, and related processes. There are two types of examinations- Type I examines the policies and procedures in place for their effectiveness to the stated objective; Type II reports on how the systems were actually used during the period of review. The SAS-70 Type II assessment is often used by hosting vendors and storage co-locations as a testament to their internal controls.

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. See Flatbed Scanner.

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 connection between computers and peripherals, such as hard disks, CD-ROM drives, and scanners. SCSI allows for up to 7 devices to be attached in a chain via cables.

SDLT (Super DLT): A type of backup tape that can hold up to 300 GB or 450 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, e.g., Google, Lycos, etc.

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 disks typically held 17 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.

Serial Port: See Port.

Server: Any central computer on a network that contains ESI or applications shared by multiple users of the network on their client computers. 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 ESI to requesting users, and name servers that provide information about Internet host names. See File Server.

Server Farm: A cluster of servers.

Service-Level Agreement: 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.

Session: A lasting connection, usually involving the exchange of many packets between a user or host and a server, typically implemented as a layer in a network protocol, such as telnet or FTP.

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

SHA-1 and SHA-2 (Secure Hash Algorithm): for computing a condensed representation of a message or a data file specified by FIPS PUB 180. See Hash.

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).

Single Instance Storage: The method of deduplication that is undertaken on a storage device to maximize space by eliminating multiple copies of a single file by retaining only one copy. This system of storage can occur either on a file level, or on a field level, where individual components of files are disassembled so that only unique parts are retained across an entire population and the reassembly of the original files is managed upon demand.

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-disk file space from the end of the logical record information to the end of the physical disk 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. See Cluster.

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

SMS (Short Message Service): A form of text message of up to 160 characters used to communicate with text over mobile networks.

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

Snapshot: See Bit Stream Backup.

Social Network: A group of people that use the Internet to share and communicate, either professionally or personally, in a public setting typically based on a specific theme or interest. For example, Facebook is a popular Social Network that allows people to connect to friends and acquaintances anywhere in the world in order to share personal updates, pictures, and experiences.

Social Media: Internet applications which permit individuals or organizations to publicly and interactively share and communicate information of generally a personal or informal nature.

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.

Software application: See Application and Software.

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

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

Spoliation: Spoliation is the destruction of records or properties, such as metadata, that 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.

SPP (Standard Parallel Port): See Port.

Spyware: A data collection program that secretly gathers information about the user and relays it to advertisers or other interested parties. Adware usually displays banners or unwanted pop-up windows, but often includes spyware as well. See Malware.

SQL (Structured Query Language): A database computer language used to manage the data in relational databases. 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).

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 that specifies the data encoding of a document's format and content. Has been virtually replaced by XML.

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 ESI. The term usually refers to mass storage devices, such as disk and tape drives.

Storage Media: See Magnetic or Optical Storage Media.

Streaming Indexing: Real-time or near real-time, indexing of data as it being moved from one storage medium to another.

Structured Data: Data stored in a structured format, such as databases or data sets. Contrast to Unstructured Data.

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. See Coding.

Subtractive Colors: Colors formed by white light minus the color absorbed by the object. 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 that 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: (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.

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

System Files: Files allowing computer systems to run; non-user-created files.

System Metadata: See File System Metadata.

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 ESI on a magnetic tape. Tape drives are sometimes used to backup large quantities of ESI due to their large capacity and cheap cost relative to other 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 ESI into categories and subcategories, sometimes displayed in a site map. Used in information retrieval to find documents that are related to a query by identifying other documents in the same category.

TCP/IP (Transmission Control Protocol/Internet Protocol): The first two networking protocols defined; enable the transfer of data upon which the basic workings of the features of the Internet operate. See Port.

Telnet (Telecommunications Network): A protocol for logging onto remote computers from anywhere on the Internet.

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

Temporary ("Temp") File: Contemporaneous files created by applications and stored on a computer for temporary use only, created to enable the processor of the computer to quickly pull back and assemble data for currently active files.

Terabyte: 1,024 gigabytes (approximately one trillion bytes). See Byte.

Text Mining: The application of data mining (knowledge discovery in databases) to unstructured textual data. Text mining usually involves structuring the input text (often parsing, along with application of some derived linguistic features and removal of others, and ultimate insertion into a database), deriving patterns within the data, and evaluating and interpreting the output, providing such ranking results as relevance, novelty, and interestingness. Also referred to as "Text Data Mining." See Data Mining.

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

Thin Client: A computer or software program which relies on a central server for processing and application resources and is used mainly for output and input of user information or commands. See Client.

Thread: A series of technologically related communications, usually 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 emails or chats, where multiple conversation threads may exist simultaneously. See Email String.

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 a standalone or part of a desktop publishing or graphics program. Thumbnails 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): A 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 1980s.

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.

Time Zone Normalization: See Nomalization.

Toggle: A switch (which may be physical or virtualized on a screen) that is either on or off, and reverses to the opposite when selected.

Tone Arm: A device in a computer that reads to/from a hard drive.

Toolbar: The row of graphical or text buttons 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.

TREC (Text Retrieval Conference): An on-going series of workshops co-sponsored by NIST and the U.S. Department of Defense.

Trojan: A malware program that contains another hidden program embedded inside it for the purpose of discretely delivering the second program to a computer or network without the knowledge of the user or administrator. See Malware.

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

TWAIN (Tool Kit Without An Interesting Name): A universal toolkit with standard hardware/software drivers for multi-media peripheral devices. Often used as a protocol between a computer and scanners or image capture equipment.

Twiki: A "WikiWiki" - enables simple form-based Web applications without programming, and granular access control (thought it can also operate in the classic 'no authentication' mode). Other enhancements include configuration variables, embedded searches, server-side includes, file attachments, and a plug-in API that has spawned over 150 plug-ins to link into databases, create charts, sort tables, write spreadsheets, make drawings, track Extreme Programming projects, and so on.

Typeface: A specific size and style of type within a family. There are many thousands of typefaces available for computers, ranging from modern to decorative.

UDP: A protocol allowing computers to send short messages to one another. See Port.

Ultrafiche: Microfiche that 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.

Under-Inclusive: When referring to data sets returned by some method of query, search, filter, or cull, results that are returned incomplete or too narrow. See False Negative.

Unicode: A 16-bit ISO 10646 character set accommodating many more characters than ASCII character set. Created as a standard for the uniform representation of character sets from all languages, thus allowing for easier internationalization. Unicode supports characters 2 bytes wide. Sometimes referred to as "double byte language." See www.unicode.org for more information. See Double Byte.

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, Document Family, Load File, and Message Unit.

UNIX: A software operating system designed to be used by many people at the same time (multi-user) capable of performing multiple tasks or operations at the same time (multi-tasking); common operating system for Internet servers.

Unstructured Data: Refers to free form data which either does not have a data structure or have a data structure not easily readable by a computer without the use of a specific program designed to interpret the data. Examples include word processing documents or slide presentations.

Upgrade: A newer version of hardware, software, or application

Upload: To move data from one's own location to another in any manner, such as via modem, network, serial cable, Internet connection or wireless signals; indicates that data is being transmitted to a location from a location . See Download.

URI (Uniform Resource Indicators): See URL.

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. Although there are many different formats, a URL might look like this: http://www.thesedonaconference.org/publications_html. See Address.

USB (Universal Serial Bus) Port: A socket on a computer or peripheral device into which a USB cable or device can be inserted; quickly replacing the use or need for serial and parallel ports as it provides a single, standardized way to easily connect many different devices. See also Key Drive and Port.

User-Added Metadata: Data, possibly work product, created by a user while copying, reviewing, or working with a file, including annotations and subjective coding information.

UTC: See Coordinated Universal Time.

UTF-8: A character encoding form of Unicode that represents Unicode code points with sequences of one, two, three, or four bytes. UTF-8 can encode any Unicode character. It is the most common Unicode encoding on the Web and the default encoding of XML. An important advantage of UTF-8 is that it is backward compatible with the ASCII encoding, which includes the basic Latin characters. Consequently, all electronic text in the ASCII encoding is conveniently also Unicode. This backward compatibility was a primary reason for the invention of UTF-8. See ASCII, Unicode, and UTF-16.

UTF-16: A character encoding form of Unicode that represents Unicode code points with sequences of one or two 16-bit code units. UTF-16 can encode any Unicode character. It is much less often used for data interchange than the UTF-8 encoding form. UTF-16 is commonly used in computer programming languages and programming APIs and is the encoding used internally for file names by Microsoft® Windows and NTFS. See Unicode and UTF-8.

Validate: In the context of this document, to 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. Vector images are typically smoother than raster images.

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 User-Added Metadata. See also Metadata.

Verbatim Coding: Manually extracting data from documents in a way that matches exactly as the information appears in the documents. See Coding.

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 documents/data are eliminated within a single custodial or production data set. See also Content Comparison, File-level Binary Comparison Horizontal De-duplication, Metadata Comparison, and Near De-Duplication.

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

Virtualization: Partitioning a server into multiple "virtual" servers, each capable of running an independent operating system and associated software applications as though it were a separate machine. Virtualization is particularly useful for centralized IT infrastructures to manage multiple computing environments with the same set of hardware, and for cloud computing providers to provide customized interfaces to clients without investing in separate machines, each with its own operating system.

Virus: A self-replicating program that spreads on a computer or network 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 that range from harmless to destructive and damage computers by either destroying data or overwhelming the computer's resources. See Malware.

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 Internet connection.

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

Volume Boot Sector/Record: 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. See Partition.

VMWare: A provider of virtualization software.

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. One minute of .WAV audio can use as much as 5 Megabytes of storage space.

Webmail: See Email (Electronic Mail).

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., FTP, telnet, or Internet sites). See URI and URL.

WiFi (Wireless Fidelity): Wireless networking technology using a quickly evolving range of networking standards.

Wiki: A collaborative Web site that allows visitors to add, remove, and edit content.

Wildcard Operator: A character used in text-based searching that assumes the value of any alphanumeric character or characters. Used to expand search terms and enable the retrieval of a wider range of hits.

Windows-1252: Also called ANSI, Western European, and CP1252 (Microsoft® code page 1252). A character encoding of the Latin alphabet used for most Western European languages. Windows-1252 is a superset of the ASCII and ISO 8859-1 standard character encodings. The characters that are included in Windows-1252 but that are not included in ISO 8859-1 are often the source of character interpretation and display problems in text on the Web and in electronic mail. Similar problems sometimes occur when text in the Windows-1252 encoding is converted to the UTF-8 encoding form of Unicode because UTF-8 is not wholly backward compatible with Windows-1252. The name ANSI is a misnomer resulting from historical happenstance, but it is not incorrect to use it in contexts where its meaning is readily understood. See ASCII and ISO 8859-1.

Wireless Router: A hardware device that opens access to an Internet connection or network to a secured or unsecured connection via a receiver on a computer or other piece of hardware such as a printer permitting wireless transmission. See WiFi.

Workflow: The automation of a business process, in whole or part, during which ESI or tasks are passed from one participant to another for action according to a set of procedural rules.

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: A self-replicating computer program, sending copies of itself, possibly without any user intervention. See Malware.

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

WWW (World Wide Web): A massive collection of hypertext documents accessed via the Internet using a browser. The documents, also known as Web Pages, can contain formatted text, audio and video files, and programs.

WYSIWYG (What You See Is What You Get): Display and software technology that shows on the computer screen exactly what will print.

X.25: A standard protocol for data communications.

XML: See Extensible Markup Language.

Yottabyte: 1,024 zettabytes. See Byte.

Zettabyte: –1,024 exabytes. See Byte.

Zip Drive: A removable disk storage device developed by Iomega with disk capacities of 100, 250, and 750 megabytes.

ZIP: A common file compression format that allows quick and easy storage for transmission or archiving.

Zone OCR: An add-on feature of imaging software that populates data fields by reading certain regions or zones of a document, and then placing the recognized text into the specified field.

— END —

Appendix A:
The Sedona Conference® Working Group Series &
WGS^{SM} Membership Program

The Sedona Conference® Working Group Series ("WGS^{SM}") represents the evolution of The Sedona Conference® from a forum for advanced dialogue to an open think-tank confronting some of the most challenging issues faced by our legal system today.

Working Groups in the WGS^{SM} begin with the same high caliber of participants as our regular season conferences. The total, active group, however, is limited to 30-35 instead of 60. Further, in lieu of finished papers being posted on the website in advance of the Conference, thought pieces, and other ideas are exchanged ahead of time, and the Working Group meeting becomes the opportunity to create a set of recommendations, guidelines, or other position piece designed to be of immediate benefit to the bench and bar, and to move the law forward in a reasoned and just way. Working Group output, when complete, is then put through a peer review process, including, where possible, critique at one of our regular season conferences, hopefully resulting in authoritative, meaningful, and balanced final papers for publication and distribution.

The first Working Group was convened in October 2002, and was dedicated to the development of guidelines for electronic document retention and production. The impact of its first (draft) publication— *The Sedona Principles; Best Practices Recommendations and Principles Addressing Electronic Document Production* (March 2003 version)—was immediate and substantial. *The Principles* was cited in the Advisory Committee on Civil Rules Discovery Subcommittee Report on Electronic Discovery less than a month after the publication of the "public comment" draft, and was cited in a seminal e-discovery decision of the Southern District of New York less than a month after that. As noted in the June 2003 issue of Pike & Fischer's *Digital Discovery and E-Evidence, "The Principles*...influence is already becoming evident."

The WGS^{SM} Membership Program was established to provide a vehicle to allow any interested jurist, attorney, academic or consultant to participate in Working Group activities. Membership provides access to advance drafts of Working Group output with the opportunity for early input, and to a Bulletin Board where reference materials are posted and current news and other matters of interest can be discussed. Members may also indicate their willingness to volunteer for special Project Team assignments, and a Member's Roster is included in Working Group publications. The annual cost of membership is only $295, and includes access to the Member's Only area for one Working Group; additional Working Groups can be joined for $100/Group.

We currently have active Working Groups in the areas of 1) electronic document retention and production; and 2) protective orders, confidentiality, and public access; 3) the role of economics in antitrust; 4) the intersection of the patent and antitrust laws; (5) *Markman* hearings and claim construction in patent litigation; (6) international issues in e-disclosure and privacy; and (7) Sedona Canada—electronic document production in Canada. See the "Working Group Series^{SM}" area of our website for further details on our Working Group Series and the Membership Program.

> "DIALOGUE DESIGNED TO MOVE
> THE LAW FORWARD IN A REASONED
> AND JUST WAY."

wgs^{SM}

wgs[SM]