

## **IMPORTANT NOTICE:**

**This Publication Has Been Superseded**

**See the Most Current Publication at**

[https://thesedonaconference.org/publication/The Sedona  
Conference Glossary](https://thesedonaconference.org/publication/The_Sedona_Conference_Glossary)

# THE SEDONA CONFERENCE JOURNAL®

*V o l u m e 1 5 ❖ F a l l 2 0 1 4*

## A R T I C L E S

**In Memoriam: Richard G. Braman** . . . . . Craig W. Weinlein

**Responding to the Government's Civil Investigations**. . . . . David C. Shonka

**What's the Problem with Google?** . . . . . Daniel R. Shulman

**Technology-Assisted Review: The Judicial Pioneers**  
. . . . . Paul E. Burns & Mindy M. Morton

**The Sedona Conference Commentary on Patent Damages  
& Remedies**. . . . . The Sedona Conference

**The Sedona Conference Commentary on Information Governance**  
. . . . . The Sedona Conference

**The Sedona Conference Database Principles Addressing the  
Preservation & Production of Databases & Database  
Information in Civil Litigation** . . . . . The Sedona Conference

**The Sedona Conference Best Practices Commentary on the  
Use of Search & Information Retrieval Methods in E-Discovery**  
. . . . . The Sedona Conference

**The Sedona Conference Commentary  
on Achieving Quality in the E-Discovery Process** . . . The Sedona Conference

**The Sedona Conference Glossary: E-Discovery & Digital  
Information Management (Fourth Edition)** . . . . . The Sedona Conference



ANTITRUST LAW, COMPLEX LITIGATION  
AND INTELLECTUAL PROPERTY RIGHTS

The Sedona Conference Journal® (ISSN 1530-4981) is published on an annual basis, containing selections from the preceding year's Conferences and Working Group efforts.

The Journal is available on a complementary basis to courthouses and public law libraries and by subscription to others (\$95; \$45 for Conference participants and Working Group members).

Send us an email ([info@sedonaconference.org](mailto:info@sedonaconference.org)) or call (1-602-258-4910) to order or for further information. Check our website for further information about our Conferences, Working Groups, and publications: [www.thesedonaconference.org](http://www.thesedonaconference.org).

Comments (strongly encouraged) and requests to reproduce all or portions of this issue should be directed to:

The Sedona Conference, 5150 North 16th Street, Suite A-215,  
Phoenix, AZ 85016 or call 1-602-258-4910; fax 602-258-2499; email [info@sedonaconference.org](mailto:info@sedonaconference.org).

*The Sedona Conference Journal*® designed by MargoBDesign.com – [mbraman@sedona.net](mailto:mbraman@sedona.net)

Cite items in this volume to “15 Sedona Conf. J. \_\_\_\_\_ (2014).”

Copyright 2014, The Sedona Conference.  
All Rights Reserved.

THE SEDONA  
CONFERENCE GLOSSARY:  
E-DISCOVERY & DIGITAL INFORMATION  
MANAGEMENT\* (FOURTH EDITION)

---

*A Project of The Sedona Conference Working Group on  
Electronic Document Retention & Production (WG1).*

*Editors:*

Sherry B. Harris, Crowley Law Office  
Paul H. McVoy, Milberg LLP

*RFP+ eDiscovery Vendor Panel*

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

*Contributing Editor:*

Matt K. Mulder, KPMG

Thanks go to all who participated in the dialogue that led to this Commentary. We thank all of our Working Group Series Sustaining and Annual Sponsors, whose support is essential to our ability to develop Working Group Series publications. For a listing of our sponsors just click on the “Sponsors” Navigation bar on the homepage of our website.

The opinions expressed in this publication, unless otherwise attributed, represent consensus views of the members of The Sedona Conference Working Group 1. They do not necessarily represent the views of any of the individual participants 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.\*

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

The Sedona Conference acknowledges the contributions of all the Working Group 1 and RFP+ eDiscovery Vendor Panel members who reviewed and commented on drafts of this Fourth Edition, and thanks all of our Working Group Series Sponsors, whose support is essential to the ability of The Sedona Conference to develop Working Group Series publications such as this. For a complete listing of our Sponsors, click on the “Sponsors” navigation bar on the home page of our website at <http://www.thosedonaconference.org>.

As with all of our publications, your comments are welcome. Please forward them to me at [cww@sedonaconference.org](mailto:cww@sedonaconference.org).

Craig Weinlein  
Executive Director  
The Sedona Conference  
Phoenix, AZ  
USA  
April 2014

**30(b)(6):** A short hand reference to Rule 30(b)(6) of the Federal Rules of Civil Procedure, under which 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 deposition notice. Sometimes the 30(b)(6) topics concern the discovery process itself, including procedures for preservation, collection, chain of custody, processing, review, and production.

**Ablate:** To burn laser-readable “pits” into the recorded layer of optical disks, DVD-ROMs and CD-ROMs.

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

**Access Control List (ACL):** A security group comprised of individual users or user groups that is used to grant similar permissions to a program, database or other security controlled environment.

**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 Record.

**Address:** A structured format for identifying the specific location or routing detail for information on a network or the Internet. These include email addresses Simple Mail Transfer Protocol (SMTP), Internet Protocol (IP) addresses and Uniform Resource Locators (URLs), commonly known as Web addresses.

**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, <http://www.aiim.org>, that focuses on Enterprise Content Management (ECM).

**Algorithm:** With regard to electronic discovery, computer script that is designed to analyze data patterns using mathematical formulas, and is commonly used to group or find similar documents based on common mathematical scores.

**Alphanumeric:** Characters composed of letters, numbers (and sometimes non-control characters, such as @, #, \$). Excludes control characters.

**Ambient Data:** See Data.

**American National Standards Institute (ANSI):** <http://www.ansi.org>—a private, non-profit organization that administers and coordinates the U.S. voluntary standardization and conformity assessment system.

**American Standard Code for Information Interchange (ASCII, 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.

**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.

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

**Applet:** A program designed as an add-on to another program, allowing greater functionality for a specific purpose other than for what the original program was designed.

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

**Application:** Software that is programmed for one or more specific uses or purposes. The term is commonly used in place of “program” or “software.” Applications may be designed for individual users, for example a word processing program or for multiple users, as in an accounting application used by many users at the same time.

**Application Programming Interface (API):** The specifications designed into a program that allows interaction with other programs. See Mail Application Programming Interface (MAPI).

**Application Service Provider (ASP):** An Internet-based organization hosting applications on its own servers within its own facilities. Customers license the application and access it through a browser over the Internet or via some other network. See Software as a Solution (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, <http://www.arma.org>.

**Artificial Intelligence (AI):** A subfield of computer science focused on the development of intelligence in machines so that the machines can react and adapt to their environment and the unknown. 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.

**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.

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

**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 (or Document Family), Message Unit and Unitization.

**Attribute:** A specific property of a file such as location, size or type. The term attribute is sometimes used synonymously with “data element” or “property.”

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

**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.

**Auto-Delete:** The use of technology to run predefined rules at scheduled intervals to delete or otherwise manage electronically stored information. May also be referred to as a janitor program or system cleanup.

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

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

**Backup:** The process of creating a copy of active data as a precaution against the loss or damage of the original data. The process is usually automated on a regular schedule, which can include the automatic expiration of older versions. The term is also used to refer to the Electronically Stored Information itself, as in, “a backup of the email server exists.” Backups can be made to any type of storage, including portable media, CDs, DVDs, data tapes or hard drives – also known as a full backup. See Differential Backup and Incremental Backup.

**Backup Data:** A copy of Electronically Stored Information that serves as a source for recovery in the event of a system problem or disaster. See Backup.

**Backup Tape:** Magnetic tape used to store copies of Electronically Stored Information, for use when restoration or recovery is required. The creation of backup tapes is made using any of a number of specific software programs and usually involves varying degrees of compression.

**Backup Tape Rotation or Recycling:** The process whereby an organization’s backups are overwritten with new data, usually on an automated schedule which should be determined by IT in consultation with records management and legal. 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.

**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.

**Basic Input Output System (BIOS):** 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.

**Batch File:** A set of commands written for a specific program to complete a discrete series of actions, for example, renaming a series of files en masse.

**Batch Processing:** The processing of multiple sets of Electronically Stored Information at one time. See Processing Data.

**Bates Number:** Sequential numbering system used to identify individual pages of documents where each page or file is assigned a unique number. Often used in conjunction with a suffix or prefix to identify a producing party, the litigation or other relevant information. See Beginning Document Number and Production Number.

**Bayesian Search:** An advanced search that utilizes the statistical approach developed by 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. For a more thorough discussion, see The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery, available for download at <https://thesedonaconference.org/download-pub/3669>.

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

**Big Data:** Enormous volumes of data, often distributed and loosely structured, that may be challenging to process with traditional technology solutions.

**Bibliographic 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 Indexing and Coding.

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

**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 file format that contains information on the placement and color of individual bits used to convey images composed of individual bits (pixels), for which the system file extension is .bmp.

**Bit Stream Backup:** A sector-by-sector/bit-by-bit copy of a hard drive; an exact copy of a hard drive, preserving all latent data in addition to the files and directory structures. See Forensic Copy.

Cited: *Nucor Corp. v. Bell*, 2008 WL 4442571, D.S.C., January 11, 2008 (C/A No. 2:06-CV-02972-DCN.), quoting The Sedona Conference *Glossary: E-Discovery & Digital Information Management*, May 2005 First Edition (“A Bit Stream Back-up is an exact copy of a hard drive, preserving all latent data in addition to the files and directory structures. Bit Stream Back-up may be created using applications such as EnCase.”).

Cited: *U.S. v. Saboonchi*, 2014 WL 1364765 (D. Mar. Apr. 7, 2014).

**Bitonal:** A bitonal image uses only black and white.

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

**Bits Per Second (BPS):** A measurement of the rate of data transfer. See Bandwidth.

**Blowback:** The term for printing Electronically Stored Information to hardcopy.

**BMP:** See Bitmap.

**Bookmark:** A link to another location, either within the current file or location or to an external location, like a specific address on the internet.

**Boolean Search:** Boolean searches use keywords and logical operators such as “and,” “or,” and “not” to include or exclude terms from a search, and thus produce broader or narrower search results. See Natural Language Search.

**Boot Sector/Record:** See Master Boot Sector/Record and Volume Boot Sector/Record.

**Broadband:** Commonly used in the context of high bandwidth Internet access made available through a variety of quickly evolving technologies.

**Brontobyte:** 1,024 yottabytes. See Byte.

**Browser:** An application used to view and navigate the World Wide Web and other Internet resources.

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

**Burn:** The process of moving or copying data to portable media such as a CD or DVD.

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

**Byte (Binary Term):** A basic measurement of most computer data consisting 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, temporary, high speed storage location that can be used to store frequently-used data for quick user access, allowing applications to run more quickly.

**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.

**Catalog:** See Index.

**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 bitonal images.

**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.

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

**Central Processing Unit (CPU):** 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.

**Certificate:** An electronic affidavit vouching for the identity of the transmitter. See Digital Certificate, Digital Signature, Public Key Infrastructure (PKI) Digital Signature.

**Chain of Custody:** Documentation regarding the possession, movement, handling and location of evidence from the time it is identified to the time it is presented in court or otherwise transferred or submitted; necessary both to establish admissibility and authenticity; important to help mitigate risk of spoliation claims.

**Checksum:** A value calculated on a set of data as a means of verifying its authenticity to a copy of the same set of data, usually used to ensure data was not corrupted during storage or transmission.

**Child:** See Document.

**Clawback Agreement:** An agreement outlining procedures to be followed if documents or Electronically Stored Information are inadvertently produced; typically used to protect against the waiver of privilege.

**Client:** (1) In a network, a computer that can obtain information and access applications on a server; (2) an application on a hard drive that relies on a server to perform some operations. See Thin 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 in a computer's memory that temporarily stores information copied or cut from a document or file.

**Cloud Computing:** “[A] model for enabling ubiquitous, 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/publications/PubsSPs.html#800-145> (last visited January 29, 2013). For further discussion see the cited NIST publication SP800-145.pdf.

**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.

**Coding:** Automated or human process by which specific information is captured from documents. Coding may be structured (limited to the selection of one of a finite number of choices) or unstructured (a narrative comment about a document). See Indexing, Verbatim Coding, Bibliographic Coding, Level Coding and Subjective Coding.

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

**Comma Separated Value (CSV):** A text file used for the transmission of fielded data that separates data fields with a comma and typically encloses data in quotation marks.

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

**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.

**Compound Document:** A file that contains multiple files, often from different applications, by embedding objects or linked data; multiple elements may be included, such as images, text, animation or hypertext. See Container File and OLE.

**Compression:** The reduction in the size of a source file or files with the use of variety of algorithms, depending on the software being used. Algorithms approach the task in a variety of ways, generally eliminates redundant information or by predicting where changes are likely to occur.

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

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

**Computer Aided or Assisted Review:** See Technology-Assisted Review.

**Computer Client or Client:** A computer or program 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. See Thin Client.

**Computer Forensics:** 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 Forensics and Forensic Copy.

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

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

**Concatenate:** Generally, to add by linking or joining to form a chain or series; the process of linking two or more databases of similar structure to enable the user to search, use, or reference them as if they were a single database.

**Concept Search:** The method of search that uses word meanings and ideas, without the presence of a particular word or phrase, to locate Electronically Stored Information related to a desired concept. 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.

**Conceptual Analytics:** Using one or more of a number of mathematical algorithms or linguistic methodologies to analyze unstructured data by themes and ideas contained within the documents enabling the grouping or searching of documents or other unstructured data by their common themes or ideas.

**Container File:** A compressed file containing multiple files; used to minimize the size of the original files for storage and/or transporting. Examples include .zip, .pst and .nsf files. The file must be ripped or decompressed to determine volume, size, record count, etc. and to be processed for litigation review and production. Also see Decompression and Rip.

Cited: *Country Vintner of North Carolina, LLC v. E. & J. Gallo Winery, Inc.*, 718 F.3d 249 (4th Cir. Apr. 29, 2013).

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

**Contextual Search:** 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.

**Cookie:** A text file containing tracking information such as dates and times of website visits, deposited by a website onto a user’s computer or mobile device. The text file is accessed each time the website 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 websites for them, including the personalization of advertising appearing on the websites.

**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 Normalization.

**Corrupted File:** A file that has become damaged in some way, such as by a virus or by software or hardware failure, so that it is partially or completely unreadable by a computer.

**Characters Per Inch (CPI):** A description of the number of characters that are contained in an inch of backup tape.

**Cross-Custodian De-Duplication:** The suppression or removal of exact copies of files across multiple custodians for the purposes of minimizing the amount of data for review and/or production. See Case De-Duplication and De-Duplication.

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

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

**Custodian:** See Record Custodian and Record Owner.

**Custodian De-Duplication:** The removal or suppression of exact copies of a file found within a single custodian's data for the purposes of minimizing the amount of data for review and/or production. Also known as Vertical De-duplication. See De-Duplication.

**Customer Relationship Management (CRM) Application:** Computer program that helps manage communications with client and contains contact information.

**Cyclical Redundancy Checking (CRC):** 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.

**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.

**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. See Active Data and Latent Data.

**Data Categorization:** The categorization and sorting of Electronically Stored Information – 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 Cell:** An individual field of an individual record. For example, in a table containing information about all of a company's employees, information about employee Joe Smith is

stored in a single record, and information about his social security number is stored in an individual Data Cell. See Field.

**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 1970's.

**Data Extraction:** The process of parsing data, including the text of the file, from any electronic documents into separate metadata fields such as Date Created and Date Last Accessed.

**Data Field:** See Field.

**Data Filtering:** The process of identifying data based on specified parameters, such as date range, author(s), and/or keyword search terms, often used to segregate data for inclusion or exclusion in the document culling or review workflow.

**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. Also used by parties to refer to production specifications during the exchange of data during discovery.

**Data Harvesting:** See Harvesting.

**Data Map:** A document or visual representation that records the physical or network location and format of an organization's data. Information about the data can include where the data is stored, physically and virtually, in what format it is stored, backup procedures in place, how the Electronically Stored Information moves and is used throughout the organization, information about accessibility of the Electronically Stored Information, retention and lifecycle management practices and policies, and identity of records custodians.

**Data Mining:** The process of knowledge discovery in databases (structured data); often techniques for extracting information, summaries or reports from databases and data sets. In the context of electronic discovery, this term often refers to the processes used to analyze a collection of Electronically Stored Information to extract evidence for production or presentation in an investigation or in litigation. See 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 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 from a prior version. The most common method of verification is hash coding by using industry accepted algorithms such as MD5, SHA1 or SHA2. See Digital Fingerprint, File Level Binary Comparison, and Hash Coding.

**Database:** 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. The collection of information is organized into a predefined formatted structure and usually organized into fields of data that comprise individual records which are further grouped into data tables. 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. For further discussion, see *The Sedona Conference Database Principles*, available for download at <http://www.thesedonaconference.org/download-pub/426>.

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

**Date/Time Normalization:** See Normalization.

**Date Created:** A common metadata field that contains the date a file was created on or moved to and the media where it currently resides.

**Date Last Accessed:** A common metadata field that contains the date a file was last accessed, meaning last opened or moved or even copied depending on the technology used to copy.

**Date Last Modified:** A common metadata field that contains the date a file was last changed either by a modification to the content or format, printed or changed by the automatic running of any macros that are executed upon the file being opened. The date last modified field does not normally reflect a change to a file's storage location or when the file was opened and read, and is thus often used as an electronic file date control field for discovery purposes.

**Date Sent:** A common metadata field that contains the date on which an email was sent.

**Date Received:** A common metadata field that contains the date on which an email was received.

**Daubert or Daubert Challenge:** *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, at 593-94 (1993), addresses the admission of scientific expert testimony to ensure that the testimony is reliable before considered for admission pursuant to Rule 702. The court assesses the testimony by analyzing the methodology and applicability of the expert's approach. Faced with a proffer of expert scientific testimony, the trial judge must determine first, pursuant to Rule 104(a), whether the expert is proposing to testify to (1) scientific

knowledge that (2) will assist the trier of fact to understand or determine a fact at issue. This involves preliminary assessment of whether the reasoning or methodology is scientifically valid and whether it can be applied to the facts at issue. Daubert suggests an open approach and provides a list of four potential factors: (1) whether the theory can be or has been tested; (2) whether the theory has been subjected to peer review or publication; (3) known or potential rate of error of that particular technique and the existence and maintenance of standards controlling the technique's operation; and (4) consideration of general acceptance within the scientific community.

**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-dupe”):** The process of comparing electronic files or records based on their characteristics and removing, suppressing or marking exact duplicate files or records within the data set for the purposes of minimizing the amount of data for review and production. De-duplication is typically achieved by calculating a file or record's hash value using a mathematical algorithm. De-duplication can be selective, depending on the agreed-upon criteria. See 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, Near Duplicates.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

**Defragment (“defrag”):** Use of a computer utility to reorganize files so they are more physically contiguous on a hard drive or other storage medium, when 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 Fragmentation.

**Deleted Data:** Information that is no longer readily accessible to a computer user due to the intentional or automatic deletion of the data. 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.

**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 reuse 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 reuse 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; and (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 in order to remove files that are generally accepted to be system generated and have no substantive value in most instances. See NIST List.

**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.

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

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

**Differential Backup:** A method of backing up data that backs up data that is new or has been changed from that last full backup.

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

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

**Digital Audio Tape (DAT):** 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.

**Digital Certificate:** Electronic records that contain unique secure values used to decrypt information, especially information sent over a public network like the Internet. See Certificate, Digital Signature and Public Key Infrastructure (PKI) Digital Signature.

**Digital Evidence Bag (DEB):** A container file format used 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.

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

**Digital Linear Tape (DLT):** A type of magnetic computer tape used to copy data from an active system for purposes of archiving or disaster recovery.

**Digital Millennium Copyright Act (DMCA):** United States copyright law enacted 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.

**Digital Rights Management (DRM):** A program that controls access to, movement or duplication of protected data.

**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 Public Key Infrastructure (PKI) Digital Signature.

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

**Digital Video Disk or Digital Versatile Disk (DVD):** A plastic disk, like a CD, on which data can be optically 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.

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

**Directory:** The organizational structure of a computer's file storage, usually arranged in a hierarchical series of folders and subfolders. Often simulated as a file folder tree.

**Disaster Recovery Tapes:** Portable magnetic storage media used to store data for backup purposes. See Backup Data, Backup Tape.

**Discovery:** The process of identifying, locating, preserving, securing, collecting, preparing, reviewing and producing facts, information and materials for the purpose of producing/obtaining evidence for utilization in the legal process. There are several ways to conduct discovery, the most common of which are interrogatories, requests for production of documents and depositions. See Electronic Discovery.

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

**Disk Mirroring:** The ongoing process of making an exact copy of information from one location to another in real time and often used to protect data from a catastrophic hard disk failure or for long term data storage. See Mirror Image and Mirroring.

**Disk Partition:** A discrete section of a computer's hard drive that has been virtually separated from one or more other partitions on the same drive.

**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 websites). 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).

**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 Attachment, Load File, Message Unit, and Unitization – Physical and Logical.

Cited: *Abu Dhabi Commercial Bank v. Morgan Stanley & Co. Inc.*, 2011 WL 3738979 (S.D.N.Y., August 18, 2011).

**Document Date:** Generally, the term used to describe the date the document was last modified or put in final form; applies equally to paper and electronic files. See Date Last Modified, Date Created, Date Last Accessed, Date Sent, Date Received.

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

**Document Type or Doc Type:** A bibliographic coding field that captures the general classification of a document, i.e., whether the document is correspondence, memo, report, article and others.

**DoD 5015:** Department of Defense standard addressing records management.

**Domain:** A group of servers and computers connected via a network and administered centrally with common rules and permissions.

**DOS:** See Microsoft-Disk Operating System (MS-DOS).

**Double Byte Language:** See Unicode.

**Download:** To move data from a remote location to a local computer or network, usually over a network or the Internet; also used to indicate that data is being transmitted from a location to a location. See Upload.

**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 files by dragging them with the mouse and dropping them in another place.

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

**Driver:** A computer program that controls various hardware devices such as the keyboard, mouse or monitor and makes them operable with the 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.

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

**Dynamic Random Access Memory (DRAM):** 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.

**Early Data Assessment:** The process of separating possibly relevant Electronically Stored Information from non-relevant Electronically Stored Information using both computer techniques, such as date filtering or advanced analytics, and human assisted logical determinations at the beginning of a case. This process may be used to reduce the volume of data collected for processing and review. Also known as Early Case Assessment (ECA).

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

**Electronic Discovery (E-Discovery):** The process of identifying, locating, preserving, collecting, preparing, reviewing, and producing Electronically Stored Information (ESI) in the context of the legal process. See Discovery.

Cited: *Gordon v. Kaleida Health*, 2013 WL 2250579 (W.D.N.Y. May 21, 2013).

Cited: *Hinterberger v. Catholic Health System, Inc.*, 2013 WL 2250603 (W.D.N.Y. May 21, 2013).

**Electronic Document Management:** The process of using a computer program to manage individual unstructured files, either those created electronically or scanned to digital form from paper. See Information Lifecycle Management.

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

**Electronic File Processing:** See Processing Data.

**Electronic Image:** An individual page or pages of an electronic document that has been converted into a static format, for example PDF or TIFF. See PDF and TIFF.

**Electronic Record:** Information recorded in a form that requires a computer or other machine to process it.

**Electronically Stored Information (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).

**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 Gmail or Yahoo Mail. See Email Message.

**Email Address:** Unique value given to individual user accounts on a domain used to route email messages to the correct email recipient most often formatted as follows: user-ID@domain-name. See Email Message.

**Email Archiving:** A systematic approach to retaining and indexing email messages to provide centralized search and retrieval capabilities. See Journaling.

**Email Client:** See Email (Electronic Mail).

**Email Message:** A file created or received via an electronic mail system. Any attachments that may be transmitted with the email message are not part of the email message but are part of the Message Unit and Document Family.

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

**Email Store:** File or database containing individual email messages. See Container File, Message Unit, OST, PST, and NSF.

**Embedded Object:** A file or piece of a file that is copied into another file, often retaining the utility of the original file’s application; for example, a part of a spreadsheet embedded into a word processing document that still allows for editing and calculations after being embedded. See Compound Document.

**EML:** File extension of a generic email message file.

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

**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 analog format into digital or raw digital data into compressed format.

**Encryption:** A procedure that renders the contents of a message or file unreadable to anyone not authorized to read it; used to protect Electronically Stored Information being stored or transferred from one location to another.

**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#:** A common metadata field that contains the bates number of the last page of a document.

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

**Enhanced Parallel Port (EPP):** See Port.

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

**Enhanced Titles:** A bibliographic coding field that captures a meaningful/descriptive title for a document based on a reading of the document as opposed to a verbatim title lifted as it appears on the face of the document. See Verbatim Coding.

**Enterprise Architecture:** Framework of information systems and processes integrated across an organization. See Information Technology Infrastructure.

**Enterprise Content Management (ECM):** Management of an organization's unstructured Electronically Stored Information, regardless of where it exists, throughout the entire lifecycle of the Electronically Stored Information.

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

**Erasable Optical Disk:** A type of optical disk that can be erased and new Electronically Stored Information added; most optical disks are read only.

**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:** The process of saving data or a subset of data in a format that can be used or imported by another system.

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

**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 software coding language 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:** The portion of an intranet site that is accessible by users outside of a company or organization hosting the intranet. This type of access is often utilized in cases of joint defense, joint venture, and vendor client relationships.

**False Negative:** A result from a search that is not correct because it fails to indicate a match or hit where one exists.

**False Positive:** A result from a search that is not correct because it indicates a match or hit where there is none.

**Fast Mode Parallel Port:** See Port.

**Federal Information Processing Standards (FIPS):** 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.

**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 file or data table used to record an individual piece of standardized data, such as the author of a document, a recipient or the date of a document.

**Field Mapping:** The process of normalizing data to the structure of an existing database for purposes of loading the data so that data loads to the correct field after validating the data type is the same, for example mapping the data from a field called Date to an existing field in a database named DocDate.

**Field Separator or Field Delimiter:** A character in a text delimited file that separates the fields in an individual record. For example, the CSV format uses a comma as the field separator. See Text Delimited File.

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

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

**File Compression:** See Compression.

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

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

**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 Data Verification, De-Duplication, Digital Fingerprint, 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:** A computer that serves as a storage location for files on a network. File servers may be employed to store Electronically Stored Information, such as email, financial data or word processing information or to backup the network. See Server.

**File Sharing:** Providing access to files or programs to multiple users on a network.

**File Signature:** See Digital Signature.

**File Slack:** See Slack/Slack Space.

**File System:** The means by which an operating system or program organizes and keeps track of Electronically Stored Information in terms of logical structures and software routines to control access to stored Electronically Stored Information, including the structure in which the files are named, stored, and organized. The file system also tracks data when a user copies, moves or deletes a file or subdirectory.

**File Table:** See Master File Table (MFT).

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

**File Transfer Protocol (FTP):** 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.

**Filename:** The name used to identify a specific file in order to differentiate it from other files, typically comprised of a series of characters a dot and a file extension (e.g., sample.doc). See File Extension, Full Path.

**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:** A small removable data storage device that uses flash memory and connects via a USB port. Can be imaged and may contain residual data. Metadata detail may not be the equivalent of Electronically Stored Information maintained in more robust storage media.

**Flash Memory:** A type of computer memory used for storage of data to a physical disk by electrical impulses.

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

**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, Mirror Image.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

**Forensics:** The scientific examination and analysis of data held on, or retrieved from, a computer 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). See Load File, 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 to store data within a particular operating system. 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 handwritten or poor quality print from forms that go into a database.

**Fragmentation:** The process by which parts of files are separately stored in different areas on a hard drive or removable disk in order to utilize available space. See Defragment.

**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 an index of all the words in a collection of Electronically Stored Information for specific characters, words, numbers and/or combinations or patterns thereof in varying degrees of complexity.

**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.”

**Geopbyte:** 1,024 brontobytes. See Byte.

**Ghost:** See Bit Stream Backup.

**GIF (Graphics Interchange Format):** A common 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 Address List (GAL):** A Microsoft Outlook directory of all Microsoft Exchange users and distribution lists to which messages can be addressed. 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).

**Global De-Deduplication:** See Case De-Duplication.

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

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

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

**Graphical User Interface (GUI, pronounced “goeey”):** An interface to a computer or device comprised of pictures and icons, rather than words and numbers by which users can interact with the device.

**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.

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

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

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

**Handwriting Recognition Software (HRS):** Software that interprets handwriting into machine readable form.

**Hard Drive:** A storage device 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 Electronically Stored Information from any media; an e-discovery vendor or specialist “harvests” Electronically Stored Information from computer hard drives, file servers, CDs, backup tapes, portable devices, and other sources for processing and loading to storage media or a database management system.

**Hash Coding:** A mathematical algorithm that calculates a unique value for a given set of data, similar to a digital fingerprint, representing the binary content of the data to assist in subsequently ensuring that data has not been modified. Common hash algorithms include MD5 and SHA. See Data Verification, Digital Fingerprint, File Level Binary Comparison.

**Head:** Devices which ride very closely to the surface of the platter on a hard drive and allow information to be read from and written to the platter.

**Header:** 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. See Message Header

**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 to the user of a computer. Some operating system files are hidden to prevent inexperienced users from inadvertently deleting or changing these essential files. See Steganography.

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

**High Technology Crime Investigation Association (HTCIA):** Computer forensics nonprofit association; resources include educational programs and Listservs.

**Hold:** See Legal Hold.

**Horizontal De-Duplication:** A way to identify Electronically Stored Information duplicated across multiple custodians or other production data sets normally by comparing hash algorithms to identify duplicates and then removing or suppressing those duplicates. See Case De-Duplication, De-Duplication.

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

**Hub:** A network device that connects multiple computers and 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 pointer in a hypertext document – usually appearing as an underlined or highlighted word or picture – upon selection sends a user to another location either within the current document or to another location accessible on the network or Internet.

**HyperText:** Text that includes hyperlinks 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.

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

**HyperText Transfer Protocol (HTTP):** 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 site. 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.

**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 actions.

**Image:** (1) To make an identical copy of a storage device, including empty sectors. Also known as creating a mirror image or mirroring the drive. See Bit Stream Backup, Forensic Copy, Mirror Image. (2) An electronic or digital picture of a document (e.g., TIFF, PDF, etc.). See Image Processing, Processing Data, Render Images.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

**Image Copy or 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, Format (noun).

**Image Key:** The name of an image and cross reference to the image's file in a document load file, often the Bates number of the page.

**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, Render Images.

**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.

**Import:** The process of bringing data into an environment or application that has been exported from another environment or application.

**Inactive Record:** 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 reactivated.

**Incremental Backup:** A method of backing up data that backs up data that is new or has been changed from that last backup of any kind, be it a full backup or the last incremental backup.

**Index/Coding Fields:** Database fields used to categorize and organize records. Often user-defined, these fields can be used for searching for and retrieving records. See Coding.

**Index:** A searchable catalog of information created 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 hard copy documents and ESI.

**Information Governance:** Information governance is the comprehensive, inter-disciplinary framework of policies, procedures, and controls used by mature organizations to maximize the value of an organization's information while minimizing associated risks by incorporating the requirements of: (1) e-discovery, (2) records & information management and (3) privacy/security into the process of making decisions about information.

**Information Lifecycle Management (ILM):** A phrase used to discuss the policies and procedures governing the management of data within an organization, from creation through destruction. See Electronic Document Management.

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

**Information Technology (IT) 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, system cleanup, and the like. See Enterprise Architecture.

**Infrastructure as a Service (IaaS):** A form of Cloud Computing whereby a third party service provider offers, on-demand, a part of its computer infrastructure remotely. Specific services may include servers, software or network equipment resources that can be provided on an as-needed basis without the purchase of the devices or the resources needed to support them. See Cloud Computing.

**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.

**Institute of Electrical and Electronic Engineers (IEEE):** 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.

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

**Intelligent Character Recognition (ICR):** 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 Handwriting Recognition Software (HRS), OCR.

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

**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.

**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 Organization for Standardization (ISO):** A worldwide federation of national standards bodies; [www.iso.org](http://www.iso.org).

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

**Internet:** A worldwide interconnected system of networks that all use the TCP/IP communications protocol and share a common address space. The Internet 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 Protocol (IP):** The principal communications protocol for data communications across the Internet.

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

**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.

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

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

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

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

**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, 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.

**Jailbreak:** A process of bypassing security restrictions of an operating system to take full control of a device.

**Java:** A platform-independent, programming language for adding animation and other actions to websites.

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

**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 items that are sent and received 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.

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

**Jump Drive:** See Flash Drive.

**Kerning:** Adjusting the spacing between two letters.

**Key Drive:** See Flash Drive.

**Key Field:** See Primary Key

**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.

**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.

**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 Electronically Stored Information 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.

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

**Legacy Data, Legacy System:** ESI which can only be accessed via software and/or hardware that has become obsolete or replaced. 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, litigation hold, or hold notice. See, The Sedona Conference Commentary on Legal Holds, September 2010, available for download at <http://www.thesedonaconference.org/download-pub/470>.

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

**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 Inc.'s image cross reference file; an ASCII delimited text file that cross references an image's Bates number to its location and file name.

**Lifecycle:** A record's lifecycle is the life span of a record from its creation or receipt to its final disposition. Usually described in three stages: (1) creation, (2) maintenance and use, and (3) archive to final disposition. See Information Lifecycle Management.

**Linear and Non-Linear Review:** Performed by humans. Linear review workflow begins at the beginning of a collection and addresses information in order until a full review of all information is complete. Non-linear review workflow is to prepare only certain portions for review, based either on the results of criteria, such as search terms, computer assisted review results or some other method, to isolate only information likely responsive. See Review.

**Linear Tape-Open (LTO):** 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.

**Link:** See Hyperlink.

**Liquid Crystal Display (LCD):** 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.

**Litigation Hold:** See Legal Hold.

**Load File:** A file that relates to a set of scanned images or electronically processed files that 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.

Cited: *Aguilar v. Immigration and Customs Enforcement Division of the United States Department of Homeland Security*, 255 F.R.D. 350, 353 (S.D.N.Y. 2008).

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

Cited: *E.E.O.C. v. SVT, LLC*, 2014 U.S. Dist. LEXIS 50114 (N.D. Ind. Apr. 10, 2014).

Cited: *Nat'l Day Laborer Org. Network v. United States Immigration & Customs Enforcement Agency*, 2011 U.S. Dist. LEXIS 11655 (S.D.N.Y. February 7, 2011).

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

**Log File:** A text file created by an electronic device or application to record activity of a server, website, computer or software program.

**Logical Entities:** An abstraction of a real-world object or concept that is both independent and unique. Conceptually, a logical entity is a noun, and its relationships to other entities are verbs. In a relational database, a logical entity is represented as a table. Attributes of the entity are in columns of the table and instances of the entity are in rows of the table. Examples of logical entities are employees of a company, products in a store's catalog, and patients' medical histories.

**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 file storage. A hard drive may contain a single or multiple volumes.

**Lossless Compression:** A method of compressing an image file, bit by bit, that results in no loss of information either during compression or extraction.

**Lossy Compression:** A method of image compression whereby storage size of image is reduced by decreasing 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.

**Magnetic/Optical Storage Media:** The physical piece of material that receives data that has been recorded using a number of different magnetic recording processes. Examples include hard drives, backup tapes, CD-ROMs, DVD-ROMs, Jaz and Zip Drives.

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

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

**Mail Application Programming Interface (MAPI):** A Windows-based software standard that enables a program to send and receive email by connecting the program to selected email servers. See API.

**Mailbox:** A term used to describe all email associated with an individual email account, whether located physically together on one server, across a server array or stored in cloud based storage.

**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 Quick Peek.

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

**Management Information Systems (MIS):** A phrase used to describe the resources, people and technology, used to manage the information of an organization.

**MAPI Mail Near-Line:** Documents stored on optical disks or compact disks that are housed in a 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 that 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.

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

**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.

**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, flash drives and hard drives. See Random Access Memory (RAM), Read Only Memory (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-Digest Algorithm 5 (MD5):** 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 Coding.

**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:** The generic term used to describe the structural information of a file that contains data about the file, as opposed to describing the content of a file. See System Generated Metadata and User Created Metadata. For a more thorough discussion, see *The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age (Second Edition)*, <https://thesedonaconference.org/download-pub/74>, and *The Sedona Conference Commentary on Ethics & Metadata*, <https://thesedonaconference.org/download-pub/3111>.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

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

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

**Microprocessor:** See CPU.

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

**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 portable wireless hub that allows users with the correct credentials to access the Internet.

**Migrated Data:** Electronically Stored Information that has been moved from one database or format to another.

**Migration:** Moving Electronically Stored Information from one computer application or platform to another; 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 another computer or when creating a preservation copy. See Forensic Copy and Image.

Cited: *White v. Graceland College Center for Professional Development & Lifelong Learning, Inc.*, 2009 WL 722056 at \*6 (D. Kan. March 18, 2009).

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

**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).

**Mount or Mounting:** The process of making off-line Electronically Stored Information available for online 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 Electronically Stored Information from mounted media to another media or to an online 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.

**MSG:** A common file format, in which emails can be saved, often associated with Microsoft Outlook email program, which preserves both the format and any associated attachment information.

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

**Multimedia Messaging Service (MMS):** A protocol of messaging that allows for the transmission of multimedia content such as pictures, video or sound over mobile networks. See Text Message.

**Multisync:** 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.

**National Institute of Standards and Technology (NIST):** A federal technology agency that works with industry to develop and apply technology measurements and standards. See NIST List.

**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, some 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, Load File.

Cited: *Akanthos Capital Management, LLC v. CompuCredit Holdings Corp.*, 2014 WL 896743 (N.D. Ga. Mar. 7, 2014).

Cited: *Covad Communications Co. v. Revonet, Inc.*, 254 F.R.D. 147 (D.D.C. 2008).

Cited: *Palar v. Blackhawk Bancorporation Inc.*, 2013 WL 1704302 (C.D. Ill. Mar. 19, 2013)

Cited: *Race Tires America, Inc. v. Hoosier Racing Tire Corp.*, 2012 WL 887593 (3d Cir. March 16, 2012).

**Native Format Review:** Review of Electronically Stored Information in its native format using either a third party viewer application capable of rendering native files in close approximation to their original application or the actual original application in which the Electronically Stored Information was created. See Review.

**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) Two or more files that are similar to a certain percentage, for example, files that are 90% similar may be identified as near duplicates; used for review to locate similar documents and review all near duplicates at one time; (2) 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. Data in near-line storage is often

stored on servers that do not have as high performance as active servers. 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. See Local-Area Network (LAN) and Wide-Area Network (WAN).

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

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

**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.

**New Technology File System (NTFS):** 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 solutions 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.

**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.

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

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

**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 Electronically Stored Information to a specific zone, often GMT, to permit advanced processing of the ESI, such as de-duplication. See 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.

**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 something.

**Object Linking and Embedding (OLE):** 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 Compound Document.

**Objective Coding:** See Coding.

**OCR (Optical Character Recognition):** A technology process that captures text from an image for the purpose of creating a parallel text file that can be associated with the image and searched in a database. OCR software evaluates scanned data for shapes it recognizes as letters or numerals. See Handwriting Recognition Software (HRS), Intelligent Character Recognition (ICR).

Cited (Glossary Third Ed.): *Country Vintner of North Carolina, LLC v. E. & J. Gallo Winery, Inc.*, 718 F.3d 249 (4th Cir. Apr. 29, 2013).

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

**Offline Data:** The storage of Electronically Stored Information outside the network in daily use (e.g., on backup tapes) that is only accessible through the offline storage system, not the network.

**Offline Storage:** Electronically Stored Information stored on removable disk (optical, compact, etc.) or magnetic tape and not accessible by the active software or server. Often used for making disaster-recovery copies of records for which retrieval is unlikely. Accessibility to offline media usually requires restoring the data back to the active server.

**Online:** Connected to a network or the Internet.

**Online Review:** The review of data on a computer, either locally on a network or via the Internet. See Review.

**Online Storage:** The storage of Electronically Stored Information as fully accessible information in daily use on the network or elsewhere.

**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):** 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 make 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 computer operating systems are UNIX, DOS, Microsoft Windows, LINUX, Mac OS and IBM z/OS. Examples of portable device operating systems are iOS, Android, Microsoft Windows and BlackBerry. 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.

Cited: *White v. Graceland College Center for Professional Development & Lifelong Learning, Inc.*, 2009 WL 722056 at \*5 (D. Kan. March 18, 2009).

**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.

**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. See Swap File.

**Parallel Port:** See Port.

**Parent:** See Document.

**Parsing:** In electronic discovery, the process by which a file is broken apart into its individual components for indexing, processing or to prepare for loading into a review database.

**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 in order that each partition can be managed separately for security or maintenance purposes. 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 text or alphanumeric string that is used to authenticate a specific user's access to a secure program, network or part of a network.

**Path:** (1) The hierarchical description of where a directory, folder or file is located on a computer or network; (2) 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 Electronically Stored Information 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.

**Peer-to-Peer or P2P:** A form of network organization that uses portions of each user's 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.

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

**Personal Address Book (PAB):** A file type to describe a Microsoft Outlook list of contacts created and maintained by an individual user for personal use.

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

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

**Personal Data (as used in the EU Data Protection Act):** Data which relate to a natural person who can be identified from the 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).

**Personal Digital Assistant (PDA):** A portable device used to perform communication and organizational tasks.

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

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

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

**Physical File Storage:** When a file is created on a computer, a sufficient number of clusters are assigned to contain the file. If the file is not large enough to completely fill the assigned clusters 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.

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

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

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

**Pixel:** A single unit of a raster image that allows a picture to be displayed on an electronic screen or computer monitor.

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

**Plasma Display:** A type of flat panel display commonly used 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.

**Platform as a Service (PaaS):** 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.

**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.

**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, allowing the document to be overwritten, but the document is not actually erased until overwritten.

**Port:** An interface between a computer and other computers or devices. 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 Universal Serial Bus (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 Transmission Control Protocol/Internet Protocol (TCP/IP) and User Datagram Protocol (UDP).

**Portable Document Format (PDF):** 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.

Cited: *Country Vintner of North Carolina, LLC v. E. & J. Gallo Winery, Inc.*, 718 F.3d 249 (4th Cir. Apr. 29, 2013).

Cited: *Saliga v. Chemtura Corp.*, 2013 WL 6182227 (D. Conn. Nov. 25, 2013).

**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).

**Precision:** When describing search results, precision is the number of documents retrieved from a search divided by the total number of documents returned. For example, in a search for documents relevant to a document request, it is the percentage of documents returned from a search that are actually relevant to the request. See The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery, available for download at <https://thesedonaconference.org/download-pub/3669>.

**Predictive Coding/Ranking:** See Technology-Assisted Review.

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

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

**Primary Key:** A unique value stored in a field or fields of a database record that is used to identify the record and, in a relational database, to link multiple tables together.

**Print On Demand (POD):** A term referring to document images stored in electronic format and available to be quickly printed.

**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:** The automated ingestion of Electronically Stored Information into a program for the purpose of extracting metadata and text; and in some cases, the creation of a static image of the source Electronically Stored Information files according to a predetermined set of specifications, in anticipation of loading to a database. Specifications can include the de-duplication of ESI, or filtering based on metadata contents such as date or email domain and specific metadata fields to be included in the final product.

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

**Production Data Set:** The universe of documents and/or Electronically Stored Information identified as responsive to document requests and not withheld on the grounds of 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 Bates Number, Beginning Document 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 Transmission Control Protocol/Internet Protocol (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 Electronically Stored Information 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 Electronically Stored Information into the paper discovery paradigm.

**Proximity Search:** A search syntax written 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.

Cited: *White v. Graceland College Center for Professional Development & Lifelong Learning, Inc.*, 2009 WL 722056 at \*5 (D. Kan. March 18, 2009).

**Public Key Infrastructure (PKI) 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.

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

**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.

**Query By Image Content (QBIC):** 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.

**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 Electronically Stored Information are made available for review or inspection before being reviewed for responsiveness, relevance, privilege, confidentiality or privacy. See Make-Available Production.

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

**Random Access Memory (RAM):** 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 to increase operating efficiency. 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 they typically have jagged edges.

**Read Only Memory (ROM):** Random memory that can be read but not written or changed. Also, hardware, usually a chip, within a computer containing programming necessary for starting 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.

**Recall:** When describing search results, recall is number of documents retrieved from a search divided by all of the responsive documents in a collection. For example, in a search for documents relevant to a document request, it is the percentage of documents returned from a search compared against all documents that should have been returned and exist in the data set. See The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery, available for download at <https://thesedonaconference.org/download-pub/3669>.

**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 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. See Information Lifecycle Management.

**Record Owner:** The physical custodian or 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. See Information Lifecycle Management.

**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:** 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. See Information Lifecycle Management.

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

**Recover, Recovery:** See Restore.

**Redaction:** A portion of an image or document is intentionally obscured or removed to prevent disclosure of a specific portion. Done to protect privileged or irrelevant portions, including highly confidential, sensitive or proprietary information.

**Redundant Array of Independent Disks (RAID):** 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.

**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; e.g., 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.

**Replication:** See Disk Mirroring.

**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 and is unintentionally left behind. 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.

**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 or piece of software to identify how it was created in order to recreate it in a new or different form. 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 or relevance to specific issues in a matter. See Native Format Review, Online Review.

**Rewritable 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.

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

**Rip:** To extract Electronically Stored Information 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 acronym of the company that developed and sells BlackBerry devices, 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.

**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 Wireless Router.

**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 (ASP), Cloud Computing.

**Sampling:** Sampling usually refers to the process of testing a database or a large volume of Electronically Stored Information 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 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.

**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.

**Search:** See Bayesian Search, Boolean Search, Compliance Search, Concept Search, Contextual Search, Full-Text Search, Fuzzy Search, Index, Keyword, Pattern Recognition, Proximity Search, Query By Image Content (QBIC), Sampling, Search Engine and Search Syntax.

**Search Engine:** A program that enables a search for keywords or phrases, such as on Web pages throughout the World Wide Web, e.g., Google, Bing, etc.

**Search Syntax:** The grammatical formatting of a search string, which is particular to the search program. Includes formatting for proximity searches, phrase searches or any other options that are supported by the search program.

**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.

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

**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 Electronically Stored Information or applications shared by multiple users of the network on their client computers; servers provide 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 Electronically Stored Information 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 File Transfer Protocol (FTP).

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

**Short Message Service (SMS):** The most common data application for text messaging communication, allows users to send text messages to phones and other mobile communication devices. See Text Message.

**Signature:** See Certificate.

**Simplex:** One-sided page(s).

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

**Single Instance Storage:** The method of de-duplication 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.

**Siri:** A personal assistant application for the Apple iPhone mobile device that allows users to search for the phone applications and Web services through voice recognition and natural language processing software.

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

**Slack/Slack Space:** The unused space that exists on a hard drive 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 Bitmap and Cluster (File).

**Small Computer System Interface (SCSI, pronounced “skuzzy”):** A common, industry standard connection type 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 Digital Linear Tape (DLT).

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

**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, and is used by entities as a public-facing presence.

**Social Media:** Internet applications which permit individuals or organizations to interactively share and communicate.

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

**Software application:** See Application, Software.

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

**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.

Cited: *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 269 F.R.D. 497(D. Md., 2010).

Cited: *Rimkus Consulting Group, Inc. v. Cammarata*, 688 F. Supp. 2d 598 (S.D. Tex., 2010).

Cited: *Quantlab Technologies Ltd. (BGI) v. Godlevsky*, 2014 WL 651944 (S.D. Tex. Feb. 19, 2014).

**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.

**Stand-Alone Computer:** A personal computer that is not connected to any other computer or network.

**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 Extensible Markup Language (XML).

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

**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.

**Storage Media:** See Magnetic/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 according to specific form and content rules as defined by each field of the database. Contrast to Unstructured Data.

**Structured Query Language (SQL):** 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).

**Subjective Coding:** Recording the judgments of a reviewer as to a document's relevancy, privilege or importance with regard to factual or legal issues in a legal matter. See Coding.

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

**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 responsible for and/or in charge of keeping a network or enterprise resource, such as a large database, operational.

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

**System-Generated Metadata:** Information about a file that is created and applied to a file by a computer process or application. Information could include the data a file was saved, printed or edited, and can include where a file was stored and how many times it has been edited. See Metadata.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

**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 Electronically Stored Information on a magnetic tape. Tape drives are sometimes used to backup large quantities of Electronically Stored Information 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 Websites and portals, a site's taxonomy is the way it organizes its Electronically Stored Information into categories and subcategories, sometimes displayed in a site map. Used in information retrieval to find documents related to a query by identifying other documents in the same category.

**Technology-Assisted Review (TAR)\*:** A process for prioritizing or coding a collection of Electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection. Some TAR methods use algorithms that determine how similar (or dissimilar) each of the remaining documents is to those coded as relevant (or non-relevant, respectively) by the subject matter experts(s), while other TAR methods derive systematic rules that emulate the expert(s) decision-making process. TAR systems generally incorporate statistical models and/or sampling techniques to guide the process and to measure overall system effectiveness.

---

\* Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Forward by John M. Facciola*, U.S. magistrate Judge, 2013 Fed. Cts. L. Rev. 7 (January 2013), <http://www.fclr.org/fclr/articles/html/2010/grossman.pdf>.

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

**Template:** 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 Delimited File:** A common format for structured data exchange whereby a text file contains fielded data where the fields are separated by a specific ASCII character and also usually contain a header line that defines the fields contained in the file. See Field Separator or Field Delimiter.

**Text Message:** A written message typically restricted to 160 characters in length that is sent among users with mobile devices. The messages can be sent via the Short Messaging Service (SMS), as well as images, video and other multimedia using the Multimedia Messaging Service (MMS).

**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.

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

**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 Electronically Stored Information storage in a central area instead of locally; 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 Flash 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.

Cited: *Akanthos Capital Management, LLC v. CompuCredit Holdings Corp.*, 2014 WL 896743 (N.D. Ga. Mar. 7, 2014).

Cited: *Country Vintner of North Carolina, LLC v. E. & J. Gallo Winery, Inc.*, 718 F.3d 249 (4th Cir. Apr. 29, 2013).

Cited: *E.E.O.C. v. SVT, LLC*, 2014 U.S. Dist. LEXIS 50114 (N.D. Ind. Apr. 10, 2014).

Cited: *Race Tires America, Inc. v. Hoosier Racing Tire Corp.*, 2012 WL 887593 (3d Cir. Mar. 16, 2012).

Cited: *Saliga v. Chemtura Corp.*, 2013 WL 6182227 (D. Conn. Nov. 25, 2013).

Cited: *In re Seroquel Products Liability Litigation*, 244 F.R.D. 650 (M.D. Fla., Aug. 21, 2007).

Cited: *Williams v. Sprint/United Management Company*, 230 F.R.D. 640 (D. Kan. 2005).

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

**TIFF Group IV:** 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 Normalization.

**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.

**Tool Kit Without An Interesting Name (TWAIN):** 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.

**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 (LAN) 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.

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

**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).

**Twiki:** A WikiWiki – Enables simple form-based Web applications without programming, and granular access control (though 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, 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.

**User Datagram Protocol (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.

**Unallocated Space:** The area of computer media, such as a hard drive, that does not contain readily 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 the ASCII character set. Created as a standard for the uniform representation of character sets from all languages. Unicode supports characters 2 bytes wide. Sometimes referred to as “double byte language.” See [www.unicode.org](http://www.unicode.org) for more information.

**Uniform Resource Locators (URL):** 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://thesedonaconference.org/publications\\_html](http://thesedonaconference.org/publications_html). See Address.

**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 Attachment, Document or Document Family, Load File, Message Unit.

Cited: *Race Tires America, Inc. v. Hoosier Racing Tire Corp.*, 2012 WL 887593 (3d Cir. Mar. 16, 2012).

**Universal Serial Bus (USB) Port:** A port 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 Flash Drive and Port.

**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 has a data structure not easily readable by a computer without the use of a specific program designed to interpret the data; created without limitations on formatting or content by the program with which it is being created. Examples include word processing documents or slide presentations.

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

**Upload:** To move data from one 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.

**User Created Metadata:** Information about a file that is created and applied to a file by a user. Information includes the addressees of an email, annotations to a document and objective coding information. See Metadata.

Cited: *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320 (Fed. Cir. Dec. 13, 2013).

**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, 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, UTF-8.

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

**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.

**Verbatim Coding:** Manually extracting information 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 Electronically Stored Information, as determined by matching hash values, are eliminated within a single custodian's data set. See Content Comparison, File level Binary Comparison Horizontal De-Duplication, Metadata Comparison, Near Duplicates.

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

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

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

**Video 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.

**Virtual Private Network (VPN):** A secure network that is constructed by using public wires to secure 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.

**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 computer. 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 re-establishment of the organization after a disaster.

**Visualization:** The process of graphically representing data.

**Voice over Internet Protocol (VoIP):** 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 a single volume may be contained on more than one disk.

**Volume Boot Sector/Record:** When a partition is formatted to create a volume of data, 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.

**WAV:** File extension name for Windows sound files.

**Webmail:** Email service that is provided through a website. See Email.

**Website:** 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.

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

**Wide Area Network (WAN):** Refers generally to a network of PCs or other devices, remote to each other, connected by electronic means, such as transmission lines. See Network.

**WiFi (Wireless Fidelity):** Wireless networking technology that allows electronic devices to connect to one another and the Internet from a shared network access point.

**Wiki:** A collaborative website 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, characters, or in some cases, words. 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, 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 Electronically Stored Information 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.

**World Wide Web (WWW):** 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 multimedia programs.

**Write Once Read Many Disks (WORM 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.

**X.25:** A standard protocol for data communications, which has largely been replaced by less complex protocols including the Internet Protocol (IP).

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

**Yottabyte:** 1,024 zettabytes. See Byte.

**Zettabyte:** 1,024 exabytes. See Byte.

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

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

**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.