



PLANET DATA
POWERED BY THE CERULEAN ENGINE

DISCOVERY MANAGEMENT GLOSSARY

Active Data: Active Data is information residing on the direct access storage media of computer systems, which is readily visible to the operating system and/or application software with which it was created and immediately accessible to users without un-deletion, modification or reconstruction.

Affidavit: is a formal sworn statement of fact, signed by the declarant (who is called the affiant or deponent) and witnessed (as to the veracity of the affiant's signature) by a taker of oaths, such as a notary public.

Alternative Dispute Resolution (ADR): extrajudicial processes such as arbitration, collaborative law, and mediation used to resolve conflict and potential conflict between and among individuals, business entities, governmental agencies, and (in the public international law context) states.

Analysis: The process of determining relevancy of paper and electronic discovery materials through evaluation based on the variables of the case.

Analytcs: A unique technology that was designed to address the challenges of unstructured information, to make computers search and process this information in a more human-like manner. Planet Data Analytic utilizes a Latent Semantic Indexing (LSI) that is a mathematically based technology. See "Latent Semantic Indexing (LSI)" below. Other Analytic applications utilize dictionary and thesaurus based technologies.

Application Programming Interface (API): is a set of routines, data structures, object classes and/or protocols provided by libraries and/or operating system services in order to support the building of applications

Archive: A copy of data on a computer drive, or on a portion of a drive, maintained for historical reference.

Archival Data: Archival Data is information that is not directly accessible to the user of a computer system but that the organization maintains for long-term storage and record keeping purposes. 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 in compressed formats.

ASCII (American Standard Code for Information Interchange): is a coding standard that can be used for interchanging information, if the information is expressed mainly by the written form of English words. It is implemented as a character-encoding scheme based on the ordering of the English alphabet. ASCII codes represent text in computers, communications equipment, and other devices that work with text. Most modern character-encoding schemes—which support many more characters than did the original—have a historical basis in ASCII.

Attachment: A memorandum, letter, spreadsheet, or any other electronic document appended to another document or email

Attorney-client Privilege: is a legal concept that protects communications between a client and his or her attorney and keeps those communications confidential.

Attribute: A characteristic that identifies it, such as type, length or location.

Audit Log / Audit Trail: A chronological sequence of audit records, each of which contains evidence directly pertaining to and resulting from the execution of a business process or system function.

Author: A person or position who originated a document.

Backup: A copy of inactive data, intended for use in the restoration of data lost to catastrophic failure of system memory. Most users backup some of their files, and many computer networks utilize automatic backup software to make regular copies of some or all of the data on the network. Some backup systems use digital audio tape (DAT) as a storage medium.

Backup Data: Backup Data is information that is not presently in use by an organization and is routinely stored separately upon portable media, to free up space and permit data recovery in the event of disaster.

Backup Tape: See Disaster Recovery Tape.

Backup Tape Recycling: Backup Tape Recycling describes the process whereby an organization's backup tapes are overwritten with new backup data, usually on a fixed schedule (e.g., the use of nightly backup tapes for each day of the week with the daily backup tape for a particular day being overwritten on the same day the following week; weekly and monthly backups being stored offsite for a specified period of time before being placed back in the rotation).

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

Batch file: Instructions defined within a file used to instruct a computer program to perform a function or series of functions.

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

Bates Numbering: A process that is commonly used as an organizational method to label and identify legal documents. During the discovery phase of litigation, a large number of documents may necessitate the use of unique identifiers for each page of each document for reference and retrieval. Such "numbering" may be solely numeric or may contain a combination of letters and numbers (alphanumeric). There is no standard method for numbering documents.

Binary: Mathematical base 2, or numbers composed of a series of zeros and ones. Since zeros and one's can be easily represented by two voltage levels on an electronic device, the binary number system is widely used in digital computing.

Bit: a binary digit, taking a value of either 0 or 1. Binary digits are a basic unit of information storage and communication in digital computing and digital information

Blu-ray Disc (also known as **Blu-ray** or **BD**) is an optical disc storage medium. Its main uses are high-definition video and data storage. The disc has the same physical dimensions as standard DVDs and CDs.

Boolean Search: The term "Boolean" refers to a system of logic developed by an early computer pioneer, George Boole. In Boolean searching, an "and" operator between two words results in a search for documents containing both of the words. An "or" operator between two words creates a search for documents containing either of the target words. A "not" operator between two words creates a search result containing the first word but excluding the second.

Burn: Slang for making (burning) a CD, DVD or Blu-Ray Disk copy of data, whether it is music, software, or other data.

Business Risk Management: A structured approach to managing uncertainty related to a threat, through a sequence of human activities including risk assessment, strategies development to manage it and mitigation of risk using managerial resources.

Byte: is a basic unit of measurement of information storage in computer science. In many computer architectures it is a unit of memory addressing. There is no standard but a byte most often consists of eight bits.

Cache: A form of high-speed **memory** used to temporarily store frequently accessed information; once the information is stored, it can be retrieved quickly from memory rather than from the hard drive.

Case De-Duplication: Retains only single copies of documents per case. For example, if an identical document resides with Mr. A, Mr. B and Mr. C, only the first occurrence of the file will be saved (Mr. A's). Contrast with custodian de-duplication and production de-duplication.

CD-ROM: Data storage medium that uses compact discs to store about 1,500 floppy discs worth of data.

Chain of Custody: refers to the chronological documentation, and/or paper trail, showing the seizure, custody, control, transfer, analysis, and disposition of evidence, physical or electronic. Because evidence can be used in court to convict persons of crimes, it must be handled in a scrupulously careful manner to avoid later allegations of tampering or misconduct which can compromise the case.

Chain of Custody Procedure: Procedure that specifies how evidence is to be moved from location to location to preserve its integrity and prove to the court that the evidence has not been altered.

Claw-back Agreement: An agreement that sets forth procedures to protect against waiver of privilege due to inadvertent production of documents or data.

Client: is an application or system that accesses a remote service on another computer system, known as a server, by way of a network.

Client Server: A client-server application is a distributed system comprising both client and server software. A client software process may initiate a communication session, while the server waits for requests from any client.

Cloud Computing: is Internet ("cloud") based development and use of computer technology ("computing")¹

Cluster: In operating systems that use a file allocation table (FAT) architecture, the smallest unit of storage space required for data written to a drive. Also called an allocation unit.

Coding: (Document Coding or Indexing): The manual extraction of key data/information from a document collection used for discovery. Such as; the Author, Recipients, Copyee's, Document Type, Document Date, and Document Characteristics. Coding for standard "bibliographic" fields is now commonly outsourced to firms where labor costs are lower than in the countries that generate the litigation in the first place. Coding of paper documents, however, will not go away until the pen is completely replaced by the computer.

Compression: A technology that reduces the size of a file. Compression programs are valuable to network users because they help save both time and bandwidth.

Computer Forensics: Computer Forensics is the use of specialized techniques for recovery, authentication, and analysis of electronic data when a case involves issues relating to reconstruction of computer usage, examination of residual data, and 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.

Cookie: Small data files written to a user's hard drive by a web server. These files contain specific information that identifies users (e.g., passwords and lists of pages visited).

Compound document: A file that combines more than one document into one by embedding objects or linked data. Data may be from different applications. The document type typically produced using word processing software, and is a regular text document intermingled with non-text elements such as spreadsheets, pictures, digital videos, digital audio, and other multimedia features. It can also be used to collect several documents into one.

Compression: A technology for storing data in fewer bits, it makes data smaller so less disk space is needed to represent the same information. Compression programs like WinZip and UNIX compress are valuable to network users because they save both time and bandwidth. Data compression is also widely used in backup utilities, spreadsheet applications, and database management systems.

Computer: a machine that manipulates data according to a list of instructions. This includes, but is not limited to, network servers, desktops, laptops, notebook computers, employees' home computers, mainframes, the PDA's of [party name] and its employees (personal digital assistants, such as Palm Pilot, Blackberry and other such handheld computing devices), digital cell phones, smart phones and pagers.

Computer Forensics: a branch of forensic science pertaining to legal evidence found in computers and digital storage mediums. Computer forensics is also known as digital forensics.

Computer Security: a branch of technology known as information security as applied to computers. The objective of computer security can include protection of information from theft or corruption, or the preservation of availability, as defined in the security policy.

Concept Search: Analyzing conceptual groups of words in a document to understand the true meaning, rather than searching only for a word (keyword).

Confidentiality: has been defined by the International Organization for Standardization (ISO) as "ensuring that information is accessible only to those authorized to have access" and is one of the cornerstones of information security.

Container file: One file that contains multiple documents and document types. Requires decompression or ripping to process.

Contextual search: Searching surrounding text to analyze the context in which a word is used.

Corporate Investigations: Criminal, regulatory, securities and/or other investigations pertaining to the activities and/or electronically stored information of one or more corporations.

Cost Sharing/Shifting: Shifting the cost or a portion of the cost of production of inaccessible electronically stored documents to the requesting party.

Culling: Removing a document prior to production or review; generally reduces the volume of data that is produced or reviewed.

Custodian: Person having administrative control of a document; for example, the data custodian of an email is the owner of the mailbox which contains the email.

Custodian De-Duplication: Culls a document if multiple copies of that document reside within the same custodian's data set. For example, if Mr. A and Mr. B each have a copy of a specific document, and Mr. C has two copies, the system will maintain one copy each for Mr. A, Mr. B, and Mr. C. Contrast with case de-duplication and production de-duplication.

Customer-Added Metadata: Data or work product created by a user while reviewing a document. For example: annotation text of a document or subjective coding information. Contrast with vendor-added metadata

DAT: Digital Audio Tape. Used as a storage medium in some backup systems.

Data: Any information stored on a computer.

Database: a structured collection of records or data that is stored in a computer system.

Database Management System (DBMS): is computer software that manages databases. In large systems, a DBMS allows users and other software to store and retrieve data in a structured way.

Data Collection: a term used to describe a process of preparing and collecting data.

Data Custodian: Person having administrative control of a document; for example, the data custodian of an email is the owner of the mailbox which contains the email.

Data Formats: The organization of information for display, storage, or printing. Data is maintained in certain common formats so that it can be used by various programs, which may only work with data in a particular format. This term is commonly used in the industry when asking another person about the state in which particular information exists. For example, "What format is it in, PDF or HTML?"

Data Hosting: A service provided for the storage and access of electronic data, images and metadata.

Data Mapping: is the process of creating data element mappings between two distinct data models.

Data Migration: the process of transferring data between storage types, formats, or computer systems.

Data Mining: is the process of extracting hidden patterns from data.

Data Set (or Dataset): is a collection of data.

Deleted File: removing or erasing a file from a computer's file system.

De-Duplication: The process of identifying (or some vendors includes actually removing) additional copies of identical documents in a document collection. There are three types of de-duplication: case, custodian, and production.

Digital Certificate: A means of providing heightened security for the access of a website or a specific document. Digital certificates are electronic records that contain keys used to decrypt information, especially information sent over a public network like the internet. Digital certificates must be applied for and granted by a Certificate Authority (CA).

Document: Any file produced by a software application.

Document Metadata: Data stored with in a document about the document. Often this data is not immediately viewable in software application used to create/edit the document, but often can be accessed via a "Properties" view. Contrast with file system metadata and email metadata. Most programs that create documents, including Microsoft SharePoint, Microsoft Word and other Microsoft Office products, save metadata with the document files. These metadata can contain the name of the person who created the file (obtained from the operating system), the name of the person who last edited the file, how many times the file has been printed, and even how many revisions have been made on the file. Other saved material, such as deleted text (saved in case of an undelete command), document comments and the like, is also commonly referred to as "metadata", and the inadvertent inclusion of this material in distributed files has sometimes led to undesirable disclosures.

De-Duplication: De-Duplication ("De-Duping") is the process of comparing electronic records based on their characteristics and removing duplicate records from the data set. The process is base on the unique HASH* algorithm. *See HASH.

Defendant: any party who is required to answer the complaint of a plaintiff or pursuer in a civil lawsuit before a court, or any party who has been formally charged or accused of violating a criminal statute.

Deleted Data: Deleted Data is data that, in the past, existed on the computer as live data and which have been deleted by the computer system or end-user activity. Deleted data remains on storage media in whole or in part until it is overwritten by ongoing usage or "wiped" with a software program specifically designed to remove deleted data. Even after the data itself has been wiped, directory entries, pointers, or other metadata relating to the deleted data may remain on the computer.

Deleted file: A file with disk space that has been designated as available for reuse. The deleted file remains intact until it has been overwritten with a new file.

Deletion: Deletion is the process whereby data is removed from active files and other data storage structures on computers and rendered inaccessible except using special data recovery tools designed to recover deleted data. Deletion occurs in several levels on modern computer

systems: (a) File level deletion: Deletion on the file level renders the file inaccessible to the operating system and normal application programs and marks the space occupied by the file's directory entry and contents as free space, available to reuse for data storage. (b) Record level deletion: Deletion on the record level occurs when a data structure, like a database table, contains multiple records; deletion at this level renders the record inaccessible to the database management system (DBMS) and usually marks the space occupied by the record as available for reuse by the DBMS, although in some cases the space is never reused until the database is compacted. Record level deletion is also characteristic of many e-mail systems. (c) Byte level deletion: Deletion at the byte level 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.

Deposition: is witness testimony given under oath and recorded for use in court at a later date.

Desktop: Usually refers to an individual PC -- a user's desktop computer.

Digital: Storing information as a string of digits – namely "1"s and "0"s.

Digital Image: is a representation of a two-dimensional image using ones and zeros (binary). Depending on whether or not the image resolution is fixed, it may be of vector or raster type. Without qualifications, the term "digital image" usually refers to raster images.

Directory, folder, catalog, or drawer is a virtual container within a digital file system, in which groups of files and other directories can be kept and organized.

Disaster Recovery Tape: Disaster Recovery Tapes are portable media used to store data that is not presently in use by an organization to free up space but still allow for disaster recovery. May also be called "Backup Tapes."

Disc (disk): It may be a floppy disk, or it may be a hard disk. Either way, it is a magnetic storage medium on which data is digitally stored. May also refer to a CD-ROM or DVD.

Disc mirroring: A method of protecting data from a catastrophic hard disk failure. As each file is stored on the hard disk, a "mirror" copy is made on a second hard disk or on a different part of the same disk.

Discovery: the pre-trial phase in a lawsuit in which each party through the law of civil procedure can request documents and other evidence from other parties or can compel the production of evidence by using a subpoena or through other discovery devices, such as requests for production of documents, and depositions. In other words, discovery includes (1) interrogatories; (2) motions or requests for production of documents; (3) requests for admissions; and (4) depositions.

Discovery Compliance: Complying with the federal, state and local regulations around discovery (e.g. Federal Rules of Civil Procedure).

Discovery Cost Distribution or Allocation: The distribution or allocation of the discovery costs incurred among multiple parties compelled to produce Hard Copy Documents and Electronically Stored information.

Discovery Response: This is a response to a discovery request.

Discovery Response Plan: A reactive or proactive plan developed to guide the activities to be taken in response to a discovery request in addition to mitigating the cost and risk.

Discovery Response Strategy: A strategic plan developed to guide the response to a request for discovery in addition to mitigating the cost and risk.

Discovery Response Team: A team of individuals assembled to coordinate and execute a Discovery Response Plan. A discovery response team may include members from legal, IT, business management and other resources from within an organization legal consulting vendors and outside counsel.

Distributed Data: Distributed Data is that information belonging to an organization which resides on portable media and non-local devices such as home computers, laptop computers, floppy disks, CD-ROMs, personal digital assistants ("PDAs"), wireless communication devices (e.g., Blackberry), zip drives, Internet repositories such as e-mail hosted by Internet service providers or portals, web pages, and the like. Distributed data also includes data held by third parties such as application service providers and business partners.

Document Imaging (Scanning): is an information technology category for systems capable of replicating documents commonly used in business. Document Imaging Systems can take many forms including microfilm, on demand printers, facsimile machines, copiers, multifunction printers, document scanners, Computer Output Microfilm (COM) and archive writers. In the last 15 years Document Imaging has been used to describe software-based computer systems that capture, store and reprint images.

Due Diligence: a term used for a number of concepts involving either the performance of an investigation of a business or person, or the performance of an act with a certain standard of care. It can be a legal obligation, but the term will more commonly apply to voluntary investigations. A common example of due diligence in various industries is the process through which a potential acquirer evaluates a target company or its assets for acquisition.

DVD: is a popular optical disc storage media format. Its main uses are video and data storage. Most DVDs are of the same dimensions as compact discs (CDs) but store more than six times as much data.

EDRM Metrics: The EDRM Metrics project is designed to provide a standard approach and generally accepted language for measuring the full range of electronic discovery activities. The Metrics project follows the electronic discovery process described in the Electronic Discovery Reference Model: identification, preservation, collection, processing, review, analysis and production. For each stage of the process, the Metrics project will offer guidelines for how to measure associated costs, time and volumes.

EDRM XML: The EDRM XML project is designed to provide a standard, generally accepted XML schema to facilitate the movement of electronically stored information (ESI) from one step of the electronic discovery process to the next, from one software program to the next and from one organization to the next.

Electronic Discovery or e-discovery: refers to discovery in civil litigation which deals with information in electronic format also referred to as Electronically Stored Information "ESI". In the legal context, electronic form is the representation of information as binary numbers. Electronic information is different from paper information because of its intangible form, volume, transience, and persistence. Also, electronic information is usually accompanied by

metadata, which is never present in paper information unless manually coded.

Electronic Mail: often abbreviated as **e-mail**, **email**, or **eMail**, is any method of creating, transmitting, or storing primarily text-based human communications with digital communications systems.

Electronically Stored Information (ESI): is any type of information that can be stored electronically, including all current types of computer-based information as well as any that might occur as a result of future changes and technological developments. Examples of ESI include E-mail messages, word processing files, voice mail messages, databases, websites and wikis. ESI is subject to electronic discovery in litigation.

Email Address: identifies a location to which e-mail messages can be delivered. An e-mail address on the modern Internet looks like, for example, jsmith@example.com and is usually read as "jsmith at example dot com".

Email Archiving: is a stand-alone IT application that works with an email server to help manage an organization's email messages. It captures and preserves all email traffic flowing into and out of the email server so it can be accessed quickly at a later date from a centrally managed location.

Email Attachment: a computer file which is sent along with an e-mail message.

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

Email Spam or Junk Email: is a subset of spam that involves nearly identical messages sent to numerous recipients by e-mail.

Embedded Metadata: Text, numbers, content, data or information that is directly or indirectly input into a Native File by a user and which is not typically visible to the user viewing the output of display of the Native File on screen or as a print-out.

Embedded Object/File: An electronic file contained within another electronic file.

Encryption: Technology that renders the contents of a file unintelligible to anyone not authorized to read it. Encryption is used to protect information as it moves from one computer to another, and is an increasingly common way of sending credit card numbers over the Internet when conducting e-commerce transactions.

ESI Processing: Capturing an electronic data image or a representation of the image, generally in native format, entering it into a computer system and processing and or manipulating it so that it can be exported into a review application.

Ethernet: A common way of networking PCs to create a LAN.

European Union Data Protection Directive 95/46/EC - Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data a European Union directive legislating protection of data pertaining to individuals. It is an important component of EU privacy and human rights law. The directive was implemented in 1995 by the European Commission.

Extranet: a private network that uses Internet protocols, network connectivity, and possibly the public telecommunication system to securely share part of an organization's information or operations with suppliers, vendors, partners, customers or other businesses.

Federal Rules of Civil Procedure: (FRCP) are rules governing civil procedure in United States district (federal) courts, that is, court procedures for civil suits.

Federal Rules of Evidence (FRE): govern the admission of facts by which parties in the federal courts of the United States may prove their cases.

File: An element of data storage in a file system. A collection of data or information that has a name, called the filename. Almost all information stored in a computer must be in a file. There are many different types of files: data files, text files, program files, directory files, and so on.

File Extension: A tag of three or four letters, preceded by a period, which identifies a data file's format or the application used to create the file. File extensions can streamline the process of locating data. For example, if one is looking for incriminating pictures stored on a computer, one might begin with the .gif and .jpg files.

File Sharing: One of the key benefits of a network is the ability to share files stored on the server among several users.

File Server: is a computer attached to a network that has the primary purpose of providing a location for the shared storage of computer files (such as documents, sound files, photographs, movies, images, databases, etc.) that can be accessed by the workstations that are attached to the computer network.

File System: The system that an operating system or program uses to organize and keep track of files. For example, a hierarchical file system is one that uses directories to organize files into a tree structure. Types of file systems include file allocation table (FAT) and Windows® NT file system (NTFS).

File System Metadata: Data that can be obtained or extracted about a file from the file system storing the file. Contrast with document metadata and email metadata.

Filename: is a special kind of string used to uniquely identify a file stored on the file system of a computer.

Filename Extension: In DOS and some other operating systems, one or several letters at the end of a filename. A suffix to the name of a computer file applied to indicate the encoding convention (file format) of its contents. Filename extensions usually follow a period (dot) and indicate the type of information stored in the file. For example, in the filename LETTER.DOC, the extension is DOC, which indicates that the file is a word processing file.

File Format: a particular way to encode information for storage in a computer file. Since a disk drive, or indeed any computer storage, can store only bits, the computer must have some way of converting information to 0s and 1s and vice-versa. There are different kinds of formats for different kinds of information. Within any format type, e.g., word processor documents, there will typically be several different formats. Sometimes these formats compete with each other.

Filtering: Electronic filtering of emails and files for privilege or by keyword, file, type, or name. Filtering removes files that do not fit the search criteria and reduces the volume of data that requires further investigation.

Firewall: an integrated collection of security measures designed to prevent unauthorized electronic access to a networked computer system. It is also a device or set of devices configured to permit, deny, encrypt, decrypt, or proxy all computer traffic between different security domains based upon a set of rules and other criteria.

Flash Drive: A portable, USB storage device that can hold between various amounts of ESI

Floppy: An increasingly rare storage medium consisting of a thin magnetic film disk housed in a protective sleeve.

Forensic Copy: A Forensic Copy is an exact bit-by-bit copy of the entire physical hard drive of a computer system, including slack and unallocated space.

Forensic Identification: the application of forensic science and technology to identify specific objects from the trace evidence often left on computer storage media; such as a hard drive.

Fragmented Data: Fragmented data is live data that has been broken up and stored in various locations on a single hard drive or disk.

FTP (File Transfer Protocol): An Internet protocol that enables you to transfer files between computers on the Internet.

GIF (Graphic interchange format): A computer compression format for pictures.

Gigabyte (GB): A gigabyte is a measure of computer data storage capacity and is a billion (1,000,000,000) bytes.

GUI (Graphical User Interface): is a type of user interface which allows people to interact with electronic devices such as computers; hand-held devices such as MP3 Players, Portable Media Players or Gaming devices; household appliances and office equipment. Examples of common contemporary operating systems include Microsoft Windows, Mac OS, Linux, BSD and Solaris.

Hard disk: A peripheral data storage device that may be found inside a desktop or laptop as in a hard drive situation. The hard disk may also be a transportable version and attached to a desktop or laptop.

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

Hart Scott Rodino Act: The Act provides that before certain mergers, tender offers or other acquisition transactions can close, both parties must file a "Notification and Report Form" with the Federal Trade Commission and the Assistant Attorney General in charge of the Antitrust Division of the Department of Justice.

Hash: is any well-defined procedure or mathematical function which converts a large, possibly variable-sized amount of data into a small datum, usually a single integer that may serve as an index into an array. The values returned by a hash function are called hash values, hash codes, hash sums, or simply hashes. Common hash algorithms include MD5 and SHA.

HTML (Hypertext Markup Language): a language that uses tags to structure text into headings, paragraphs, lists and links. It tells a Web browser how to display text and images.

Imaged Copy: A "mirror image" bit-by-bit copy of a hard drive, i.e., a complete replication of the physical drive regardless of how the drive is organized or whether the image created contains meaningful data in whole or in part. From an imaged copy of a hard drive it is possible to reconstruct the entire contents and organization of the source drive from which it was taken.

Image: In data recovery parlance, to image a hard drive is to make an identical copy of the hard drive, including empty sectors. A kin to data cloning. Also known as creating a "mirror image" or "mirroring" the drive.

Information Governance: The organizational structures and processes that ensure an accountability framework for use by IT that also support an organization's legal objectives and strategies.

Information Management: Information management is the collection and management of information from one or more sources and the distribution of that information to one or more audiences. It is largely limited to files, file maintenance, and the life cycle management of paper and electronically based files, other media and records.

Input Device: Any object which allows a user to communicate with a computer by entering information or issuing commands (e.g. keyboard, mouse or joystick).

Instant messaging (IM): is a form of real-time communication between two or more people based on typed text. The text is conveyed via devices connected over a network such as the Internet.

Internet: The interconnecting global public network made by connecting smaller shared public networks. The most well-known Internet is the Internet, the worldwide network of networks which use the TCP/IP protocol to facilitate information exchange.

Intranet: a private computer network that uses Internet technologies to securely share any part of an organization's information or operational systems with its employees.

IP address: A string of four numbers separated by periods used to represent a computer on the Internet.

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

ISO - International Organization for Standardization: an international-standard-setting body composed of representatives from various national standards organizations.

ISP (Internet Service Provider): is a company that offers its customers access to the Internet.

JPEG (Joint Photographic Experts Group) An image compression standard for photographs.

Keyword Search: A search for documents containing one or more words that are specified by a user.

Kilobyte (KB): One thousand bytes of data is 1K of data.

LAN (Local Area Network): Usually refers to a network of computers in a single building or other discrete location.

Latent Semantic Analysis (LSA): is a technique in natural language processing, in particular in vectorial semantics, of analyzing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms.

Legacy Data: Legacy Data is information in the development of which an organization may have invested significant resources and which has retained its importance, but which has been created or stored by the use of software and/or hardware that has been rendered outmoded or obsolete.

Legal Brief: refers to the research filed by an attorney or by a party to a court. The research points out the legal issue that is being raised, what the law or rule of law says about the issue, how the law should be applied, and the conclusion of the information provided. A legal brief is a written statement submitted in a trial or appellate proceeding that explains one party's side.

Legal Document Management: The policies, procedures, planning and other activities around the storage and possessing of documents that may be needed for legal matters.

Legal Hold: is a process which an organization uses to preserve all forms of relevant information when litigation is reasonably anticipated.

Legal Professional Privilege: protects all communications between a professional legal adviser (a solicitor, barrister or attorney) and his or her clients from being disclosed without the permission of the client. The privilege is that of the client and not that of the lawyer.

Litigation Management: The business activities around preparing for and/or responding to litigation.

Litigation Preparation: The strategic planning and/or activities around preparing for litigation.

Litigation Readiness Consulting: Consultative services to help guide an organization in preparation for litigation.

Litigation Response Consulting: Consultative services to help guide an organization in its response to litigation.

Litigation Support: Personnel or resources that help one or more organizations prepare for and respond to litigation or investigation.

Litigation Support Services: Services to support the preparation for and response to litigation or investigation.

LFP File: An ASCII delimited text file required for cross-reference of images to data.

Load File: A file that relates to a set of scanned images and indicates where individual pages belong together as documents. A load file may also contain data relevant to the individual

documents, such as metadata, coded data and the like. Load files must be obtained and provided in prearranged formats to ensure transfer of accurate and usable images and data.

Login: (logging or signing in) is the process by which individual access to a computer system is controlled by identification of the user using credentials provided by the user. A user can log in to a system to obtain access, and then *log out* when the access is no longer needed.

Lotus Notes: is a client-server, collaborative application developed and sold by IBM Software Group. IBM defines the software as an "integrated desktop client option for accessing business e-mail, calendars and applications on [an] IBM Lotus Domino server.

Magnetic/Optical Storage Media: Includes, but is not limited to, hard drives (also known as "hard disks"), backup tapes, CD-ROMs, DVD-ROMs, Jazz and Zip drives, and floppy discs, all used singly or in combination in, or in conjunction with, your computers and any and all backup and archive systems for the same.

Magnetic Tape: a medium for magnetic recording generally consisting of a thin magnetizable coating on a long and narrow strip of plastic. Nearly all recording tape is of this type, whether used for recording audio or video or for computer data storage.

Mailbox: An area in memory or on a storage device where email is placed. In email systems, each user has a private mailbox. When the user receives email, the mail system automatically puts it in the mailbox. The mail system allows you to scan mail that is in your mailbox, copy it to a file, delete it, print it, or forward it to another user. The mailbox format used by Microsoft Exchange® email systems is PST, while Lotus Notes® uses NSF files.

Mail Server: is also used to mean a computer acting as an MTA that is running the appropriate software.

Mail Transfer Agent (MTA): is a computer program or software agent that transfers electronic mail messages from one computer to another.

Meet and Confer - FRCP Rule 26(f): A settlement conference is a meeting between opposing sides of a lawsuit at which the parties attempt to reach a mutually agreeable resolution of their dispute without having to proceed to a trial. Such a conference may be initiated through either party, usually by the conveyance of a settlement offer; or it may be ordered by the court as a precedent (preliminary step) to holding a trial. Each party, the plaintiff and the defendant, is usually represented at the *settlement conference* by their own Counsel or attorney.

Megabyte (MB): A million bytes of data is a megabyte, or simply a meg.

Memory: Internal storage areas in the computer. The term memory identifies data storage that comes in the form of chips, and the word storage is used for memory that exists on tapes or disks. Moreover, the term memory is usually used as a short-hand for physical memory, which refers to the actual chips capable of holding data. Some computers also use virtual memory, which expands physical memory onto a hard disk. See the definitions for two types of physical memory: **RAM** and ROM.

Metadata (meta data, or sometimes meta-information) is "data about other data", of any sort in any media. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items and hierarchical levels, for example a database schema. In data processing, metadata is definitional data that provides information

about or documentation of other data managed within an application or environment. The term should be used with caution as all data is about something, and is therefore metadata.

For example, metadata would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.). Metadata may include descriptive information about the context, quality and condition, or characteristics of the data. It may be recorded with high or low granularity.

Microsoft Exchange Server: is a messaging and collaborative software product developed by Microsoft. It is part of the Microsoft Servers line of server products and is widely used by enterprises using Microsoft infrastructure solutions. Exchange's major features consist of electronic mail, calendaring, contacts and tasks; support for mobile and web-based access to information; and support for data storage.

Microsoft SQL Server: a relational database management system (RDBMS) produced by Microsoft. Its primary query languages are MS-SQL and T-SQL.

Mirroring: The duplication of data for purposes of backup or to distribute network traffic among several computers with identical data.

MIS: Management information systems.

Modem: Hardware that lets a computer talk to another computer over a phone line.

Network: A group of computers or devices that is connected together for the exchange of data and sharing of resources.

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

NSF: A file that is created by Lotus to maintain Lotus Notes Electronic Mails with attachments.

Native File: The source document, as collected from the source computer or server, before any conversion or processing of the document.

Native File Review: Reviewing ESI using the software used to create it originally. For example: using Microsoft Word in the review process to open/review a .DOC (MS Word Document format) file.

Network: a group of interconnected computers. Networks may be classified according to a wide variety of characteristics. (e.g. Local Area Network (LAN), Wide Area Network (WAN), Metropolitan Area Network (MAN), Storage Area Network (SAN), peer-to-peer network, client-server network).

Network Operating System: Software which directs the overall activity of networked computers.

Near de-duplication: The elimination of electives with "near duplicate" similarities, i.e. a document that was sent to multiple custodians.

NIST-National Institute of Standards and Technology: a measurement standards laboratory which is a non-regulatory agency of the United States Department of Commerce.

The institute's mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve quality of life.

Object Linking and Embedding (OLE): a technology that allows embedding and linking to documents and other objects developed by Microsoft.

OCR - Optical Character Recognition: is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner) into machine-editable text.

Offline: Not connected (to a network).

Ongoing Preservation Obligation: Once an organization is served with a litigation notice, all future relevant electronic communication is also subject to the legal hold.

Online / Offline: In general, "online" indicates a state of connectivity, while "offline" indicates a disconnected state.

Onsite Discovery Management: Discovery management services performed at a client's site(s). Examples: Consulting, Scanning, Coding, Electronic discovery and review services.

Operating System (OS or O/S): is an interface between hardware and applications; it is responsible for the management and coordination of activities and the sharing of the limited resources of the computer.

Overwrite: To copy new data over existing data. Overwritten data cannot be retrieved.

Parent-child Relationships: Parent-child relationship is a term used in e-discovery to describe a chain of documents that stems from a single e-mail or storage folder. These types of relationships are primarily encountered when a party is faced with a discovery request for e-mail. A "child" (i.e., an attachment) is connected to or embedded in the "parent" (i.e., an e-mail or Zip file) directly above it.

Password: a secret word or string of characters that is used for authentication, to prove identity or gain access to a resource (Example: An access code is a type of password).

Personal Computer (PC) is any general-purpose computer whose original sales price, size, and capabilities make it useful for individuals, and which is intended to be operated directly by an end user, with no intervening computer operator.

PDA (Personal digital assistant): Handheld digital organizers.

PDF: Portable Document Format - a file format developed by Adobe Systems. PDF captures formatting information from a variety of desktop publishing applications, making it possible to send formatted documents and have them appear on the recipient's monitor or printer as they were intended. To view a file in PDF format, you need Adobe Acrobat Reader, a free application distributed by Adobe Systems.

Petabyte (PB): is a unit of information or computer storage equal to one quadrillion bytes, or 1024 terabytes.

Phishing: the criminally fraudulent process of attempting to acquire sensitive information such

as usernames, passwords and credit card details by masquerading as a trustworthy entity in an electronic communication

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

Plaintiff: also known as a **claimant** or **complainant**, is the party who initiates a lawsuit (also known as an *action*) before a court. By doing so, the plaintiff seeks a legal remedy, and if successful, the court will issue judgment in favor of the plaintiff and make the appropriate court order (e.g., an order for damages).

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

Precedent: establishing a principle or rule that a court or other judicial body adopts when deciding subsequent cases with similar issues or facts.

Preservation: The process of retaining and protecting all relevant evidence from destruction or deletion.

Privacy Law: the area of law concerned with the protection and preservation of the privacy rights of individuals. Increasingly, governments and other public as well as private organizations collect vast amounts of personal information about individuals for a variety of purposes. The law of privacy regulates the type of information which may be collected and how this information may be used.

Private Area Network: A network that is connected to the Internet but is isolated from the Internet.

Privilege: a special entitlement or immunity granted by a government or other authority to a restricted group, either by birth or on a conditional basis. Example: Attorney-client Privilege or Legal Professional Privilege.

Privilege Data Set: A set of documents that are deemed responsive or relevant but are withheld on the grounds of privilege (work product or attorney-client).

Production: To electronically deliver ESI to a variety of recipients or for use in other systems.

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

Project Management: the discipline of planning, organizing and managing resources to bring about the successful completion of specific project goals and objectives.

Proximity Search: the process looks for documents where two or more separately matching term occurrences are within a specified distance, where distance is the number of intermediate words or characters.

PST File: A file that is created by Microsoft to maintain Exchange and Outlook Electronic Mails.

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

Query Languages: are computer languages used to make queries into databases and information systems.

Quality Control: The process ensuring that products or services are designed and produced to meet or exceed customer requirements. These systems are often developed in conjunction with other business and engineering disciplines using a cross-functional approach.

Quick Peek: ESI is made available to opposing party before being reviewed for privilege, confidentiality or privacy. Strict guidelines are required to prevent waiver.

RAM (Random Access Memory): The working memory of the computer into which application programs can be loaded and executed.

Raw Data: is a term for unprocessed data, it is also known as primary data.

Record Retention Policy: Policy for setting procedures around managing the lifecycle of records, from creation to maintenance to disposition.

Record Retention Schedule: A formalized plan for the management of records, identifying how long records should be kept, when they should be archived and when they can be destroyed.

Records Management, or RM, is the practice of identifying, classifying, archiving, preserving, and destroying records.

Relational Database Management System (RDBMS) is a database management system (DBMS) in which data is stored in the form of tables and the relationship among the data is also stored in the form of tables.

Repository: is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers.^[1] The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system.

Repository Hosting: A device accessed through the inter/intranet on which electronic data, images and record metadata is stored.

Reprography: is the reproduction of graphics through mechanical or electrical means, such as photography or xerography.

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

Review: Examination of potentially relevant data sets, Paper or ESI, for relevancy, privilege and confidentiality in advance of production.

ROM: Read Only Memory - the hardware in a computer that that can be read but not written to. ROM contains the programming that allows a computer to boot up each time the user turns it on, and it contains essential system programs that neither the user or the computer can erase.

Router: A piece of hardware that routes data from a local area network (LAN) to a phone line.

Rule 16: Pretrial conference - Rule 16 may provide a party with an opportunity to discuss settlement without giving the appearance of having initiated the conversation.

Rule 26: General provisions governing discovery; duty of disclosure.

Rule 34: Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes

Rule 37 (FRCP): FRCP 37(e), formerly 37(f), provides a safe harbor when data is lost or overwritten in the normal course of business.

Rule 502 (FRE)

The proposed Federal Evidence Rule 502 is intended to reduce the risk of forfeiting the attorney-client privilege or work product protection.

Sampling: Sampling usually (but not always) refers to the process of statistically testing a data set for the likelihood of relevant information. It can be a useful technique in addressing a number of issues relating to litigation, including decisions as to which repositories of data should be preserved and reviewed in a particular litigation, and determinations of the validity and effectiveness of searches or other data extraction procedures. Sampling can be useful in providing information to the court about the relative cost burden versus benefit of requiring a party to review certain electronic records.

Sandbox: A network or series of networks that are not connected to other networks.

SAS 70 (Statement on Auditing Standards No. 70): Service Organizations, commonly abbreviated as SAS 70, is an auditing statement issued by the Auditing Standards Board of the American Institute of Certified Public Accountants (AICPA), officially titled "Reports on the Processing of Transactions by Service Organizations".

Second Request: is a discovery procedure by which the Federal Trade Commission and the Antitrust Division of the Justice Department investigates mergers and acquisitions which may have anticompetitive consequences.

Under the Hart-Scott-Rodino Antitrust Improvements Act, before certain mergers, tender offers or other acquisition transactions can close, both parties to the deal must file a "Notification and Report Form" with the Federal Trade Commission (FTC) and the Assistant Attorney General in charge of the Antitrust Division.

If either the FTC or the Antitrust Division has reason to believe the merger will impede competition in a relevant market, they may request more information by way of "Request for Additional Information and Documentary Materials", more commonly referred to as a "Second Request".

Secure Data Hosting: A service provided for the secure storage and access of electronic

data, images and metadata.

Secure Sockets Layer (SSL), are cryptographic protocols that provide security and data integrity for communications over TCP/IP networks such as the Internet.

Server: Any computer on a network that contains data or applications shared by users of the network on their client PCs.

Service Level Agreement – SLA: is a part of a service contract where the level of service is formally defined. In practice, the term *SLA* is sometimes used to refer to the contracted delivery time (of the service) or performance.

Settlement: when the parties to a dispute (both disputes that are being litigated before the courts, and disputes where court action has not been started) reach an agreement as to the case, which is said to 'settle' the claim.

Simple Mail Transfer Protocol (SMTP): is an Internet standard for electronic mail (e-mail) transmission across Internet Protocol (IP) networks.

Slack: The difference in empty bytes of the space that is allocated in clusters minus the actual size of the files. Also described as the data fragments stored randomly on a hard drive during the normal operation of a computer, or the residual data left on the hard drive after new data has overwritten some of the previously stored data.

Software: is a general term used to describe a collection of computer programs, procedures and documentation that perform some tasks on a computer system.

Software Application: is any tool that functions and is operated by means of a computer, with the purpose of supporting or improving the software user's work. In other words, it is the subclass of computer software that employs the capabilities of a computer directly and thoroughly to a task that the user wishes to perform.

Spoliation: Refers to the intentional or negligent withholding, hiding, alteration or destruction of evidence relevant to a legal proceeding, and it is a criminal act in the United States under Federal and most State law.

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

Storage Device: Any device that a computer uses to store information

Storage Media: Any removable device that stores data. See magnetic or optical storage media.

Subpoena: commonly defined as a written command to a person to testify before a court or be punished.

Structured Data: Data that has a structured format, such as a database.

Structured Storage: (variously also known as *COM structured storage* or *OLE structured storage*) is a technology developed by Microsoft as part of its Windows operating system for storing hierarchical data within a single file.

System Administrator: (sys admin, sysop) a person employed to maintain and operate a

computer system and/or network. System administrators may be members of an information technology department.

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

Terabyte (TB): is a measurement term for data storage capacity. The value of a terabyte based upon a decimal radix (base 10) is defined as one trillion (short scale) bytes, or 1000 gigabytes.

TIFF (Tagged Image File Format): One of the most widely supported file formats for storing bit-mapped images. Files in TIFF format often end with a .tiff extension.

Transport Layer Security (TLS): are cryptographic protocols that provide security and data integrity for communications over TCP/IP networks such as the Internet. TLS' predecessor is Secure Sockets Layer (SSL).

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

Unicode: a computing industry standard allowing computers to consistently represent and manipulate text expressed in most of the world's writing systems, both foreign and domestic. Developed in tandem with the Universal Character Set standard and published in book form as The Unicode Standard, Unicode consists of a repertoire of more than 100,000 characters, a set of code charts for visual reference, an encoding methodology and set of standard character encodings, an enumeration of character properties such as upper and lower case, a set of reference data computer files, and a number of related items, such as character properties, rules for normalization, decomposition, collation, rendering and bidirectional display order (for the correct display of text containing both right-to-left scripts, such as Arabic or Hebrew, and left-to-right scripts).

Unitization: The assembly of individually 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 a 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 cues.

US-EU Safe Harbor: a streamlined process for US companies to comply with the EU Directive 95/46/EC on the protection of personal data. Intended for organizations within the EU or US that store customer data, the Safe Harbor Principles are designed to prevent accidental information disclosure or loss. US companies can opt into the program as long as they adhere to the 7 principles outlined in the Directive.

User: a person who uses a computer or Internet service. A user may have a user account that identifies the user by a username (also user name), screen name (also screen name).

Unstructured data (or unstructured information) refers to (usually) computerized information that either does not have a data model or has one that is not easily usable by a computer program.

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

VPN (Virtual Private Network): a computer network in which some of the links between nodes are carried by open connections or virtual circuits in some larger network (e.g., the Internet) as opposed to running across a single private network.

Web Browser: a software application which enables a user to display and interact with text, images, videos, music, games and other information typically located on a Web page at a Web site on the World Wide Web or a local area network.

Web Page or Webpage: is a resource of information that is suitable for the World Wide Web and can be accessed through a web browser. This information is usually in HTML or XHTML format, and may provide navigation to other web pages via hypertext links.

Wireless Communication: is the transfer of information over a distance without the use of electrical conductors or "wires".

World Wide Web -WWW: (commonly abbreviated as "**the Web**") is a system of interlinked hypertext documents accessed via the Internet.

World Wide Web Base Repository: A device accessed through the internet on which electronic data, images and record metadata is stored.

Extensible Markup Language – XML: is a general-purpose specification for creating custom markup languages. It is classified as an extensible language, because it allows the user to define the mark-up elements.

Zubulake: Five landmark decisions on e-discovery addressing when to shift the cost of electronic discovery to the requesting party; when a company needs to begin preserving electronic evidence and what electronic evidences must be preserved; what steps must be taken to preserve and the consequences of the failure to adequately preserve electronic evidence.