

Common Litigation Support Terms

Term	Definition
Annotation	The changes or additions made to a document using sticky notes, a highlighter, or other electronic tools.
Attachment	A memorandum, letter, spreadsheet, or any other electronic document appended to another document or email.
Boolean search	Boolean Searches use the logical operators “and”, “or” and “not” to include or exclude terms from a search.
Chain of evidence	The "sequencing" of the chain of evidence follows this order: identification and collection; analysis; storage; preservation; transportation; presentation in court; return to owner.
Concept search	Maps relationships between each word and every other word in large sets of documents and then associates words based on the context in which they are used. Two techniques can be used to perform concept searches: the use of a manually constructed thesaurus which relates certain words to others or semantic indexing, a fully automated methods to show associations among words based, in part, on statistical analysis of the occurrence of proximity of certain words to others.
De-duplication	The process of identifying (and/or removing) additional copies of identical documents in a document collection.
Discovery	The disclosure of facts, documents, electronically stored information and tangible objects by an adverse party.
eDiscovery (ED; EDD; EED)	Also called "digital discovery," "electronic digital discovery," "electronic document discovery" and "electronic evidence discovery." Discovery documents produced in electronic formats rather than hardcopy.
Electronic document	A document that has been scanned, or was originally created on a computer.
Electronically Stored Information (ESI)	Any information created, stored, or best utilized with computer technology of any type. It includes but is not limited to data; word-processing documents; spreadsheets; presentation documents; graphics; animations; images; e-mail and instant messages (including attachments); audio, video, and audiovisual recordings; voicemail stored on databases; networks; computers and computer systems; servers; archives; backup or disaster recovery systems; discs, CDs, diskettes, drives, tapes, cartridges and other storage media; printers; the Internet; personal digital assistants; handheld wireless devices; cellular telephones; pagers; fax machines; and voicemail systems.
Encryption	The coding of messages to increase security and make transmission only readable by recipients with the ability to decode only by using the same algorithms.

Term	Definition
Forensics	For electronic data, the discovery discipline that includes the physical acquisition of digital data using a methodology that satisfies evidentiary requirements of chain-of-custody and authentication. Forensics can include preserving the evidence, performing code and encryption cracking, searching and retrieving elusive data, determining if files have or have not been deleted, recovering deleted files, and determining use, including Internet, network access, printing, filing, and copying.
Full text search	Every word in the ESI is indexed into a master word list with pointers to the location within the ESI where each occurrence of the word appears.
Fuzzy search	Subjective content searching (as compared to word searching of objective data). Fuzzy Searching lets the user find documents where word matching does not have to be exact, even if the words searched are misspelled due to optical character recognition (OCR) errors. This search locates all occurrences of the search term, as well as words that are "close" in spelling to the search term.
Hash	An algorithm that creates a value to verify duplicate electronic documents. A hash mark serves as a digital thumbprint.
Index	The searchable catalog of documents created by search engine software. Also called "catalog." Index is often used as a synonym for search engine.
Keyword search	A search for documents containing one or more words that are specified by a user.
Load file	A file that relates to a set of scanned images or electronically processed files, and indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends. A load file may also contain data relevant to the individual documents, such as metadata, coded data, text, and the like. Load files must be obtained and provided in prearranged formats to ensure transfer of accurate and usable images and data.
Logical unitization	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 cues.
Mailbox	An area on a storage device where email is placed. In email systems, each user has a private mailbox. When the server receives email, the mail system automatically puts it in the appropriate mailbox.

Term	Definition
Metadata	Data typically stored electronically that describes characteristics of ESI, found in different places in different forms. Can be supplied by applications, users or the file system. Metadata can describe how, when and by whom ESI was collected, created, accessed, modified and how it is formatted. Can be altered intentionally or inadvertently. Certain metadata can be extracted when native files are processed for litigation. Some metadata, such as file dates and sizes, can easily be seen by users; other metadata can be hidden or embedded and unavailable to computer users who are not technically adept. Metadata is generally not reproduced in full form when a document is printed to paper or electronic image.
Mirror image	Used in computer forensic investigations and some electronic discovery investigations, a mirror image is a bit-by-bit copy of a computer hard drive that ensures the operating system is not altered during the forensic examination. May also be referred to as “disc mirroring,” or as a “forensic copy.”
Native application	Any application used to create and view a particular application file type.
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 MicrosoftWord document may require the MicrosoftWord application), documents may be converted to a neutral format as part of the record acquisition or archive process. “Static” formats (often called “imaged formats”), such as TIFF or PDF, are designed to retain an image of the document as it would look viewed in the original creating application but do not allow metadata to be viewed or the document information to be manipulated. In the conversion to static format, the metadata can be processed, preserved and electronically associated with the static format file. However, with technology advancements, tools and applications are becoming increasingly available to allow viewing and searching of documents in their native format, while still preserving all metadata.
Nesting	Document nesting occurs when one document is inserted within another document (i.e., an attachment is nested within an email; graphics files are nested within a Microsoft Word document).
Objective coding	Extracting information from electronic documents such as date created, author, recipient, CC, and linking each image to the information in pre-defined objective fields. In direct opposition to subjective coding where legal interpretations of data are linked to individual documents.

Term	Definition
Optical Character Recognition (OCR)	A technology process that translates and converts printed matter on an image into a format that a computer can manipulate (ASCII codes, for example) and, therefore, renders that matter text searchable. OCR software evaluates scanned data for shapes it recognizes as letters or numerals. All OCR systems include an optical scanner for reading text, and software for analyzing images. Most OCR systems use a combination of hardware (specialized circuit boards) and software to recognize characters, although some inexpensive systems operate entirely through software. Advanced OCR systems can read text in a large variety of fonts, but still have difficulty with handwritten text. OCR technology relies upon the quality of the imaged material, the conversion accuracy of the software, and the quality control process of the provider. The process is generally acknowledged to be between 80 and 99 percent accurate.
Physical unitization	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.
Proximity search	Retrieves a word only when it occurs within a specific number of lines or words of another word.
Redact	A portion of an image or document is intentionally concealed to prevent disclosure of specific portions. Often done to conceal and protect privileged portions or avoid production of irrelevant portions that may contain highly confidential, sensitive or proprietary information.
Spoliation	Spoliation is the destruction or alteration of evidence during on-going litigation or during an investigation or when either might occur sometime in the future. Failure to preserve data that may become evidence is also spoliation.
Subjective coding	The coding of a document using legal interpretation as the data that fills a field. Usually performed by paralegals or other trained legal personnel.
Tagged Image File Format (TIFF)	Graphic files that portray a single page of a file for viewing purposes with a .tif extension (in the case of Multi-page TIFFs, output images can consist of multiple pages).
Unicode	Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language (using more than one byte to represent each character, Unicode enables most written languages in the world to be represented using a single character set).