

COMMON E-DISCOVERY TERMS AND DEFINITIONS

Admissible	Evidence that is allowable in court
Analytics	The term used to refer to the various technologies used to provide multiple views into the data set
Archive	Long term repositories for the storage of records and files: zip, rar, gzip and other file types are commonly referred to as archives or containers. A single archive file commonly contains multiple individual files when are compressed and potentially password protected.
Attachment Backup	Both the action of and the result of creating a copy of data as a precaution against the loss or damage of the original data.
Backup tape	Portable media used to store copies of data that are created as a precaution against the loss or damage of the original data.
Batch Processing	The processing of a large amount of ESI in a single step
Chain-of-Custody	The purpose of the chain of custody is to track and record the movement, integrity and location of electronic evidence from collection to production.
Child Document	A file that is attached to another communication file, such as an attachment to an email embedded in a word processing document.
Clawback Agreement	This is an agreement that sets forth procedures to protect against waiver of privilege due to inadvertent production of documents or data.
Container File	A single file containing multiple documents and/or files, usually in a compressed format; e.g. zip, rar, pst. E01, AD1, DD, VHD are common container files used when collecting E-Discovery data and forensic images.
Custodian	Most often refers to the individual from whose file systems a group of records were extracted. This person is not necessarily the author of the documents.
Data Extraction	The process of parsing data from electronic documents to identify their metadata and body contents.
Data Mapping	The process of identifying and recording the location and types of ESI within an organization's network and policies and procedures related to that ESI.
De-duplication	De-duping. The process of comparing the characteristics of electronic documents to identify and/or remove duplicate records to reduce review time and increase coding consistency.
De-NIST	The process of separating computer system generated files from user-created files. This automated process utilizes a list of file extensions developed by the National Institute of Standards and Technology (NIST).
Discovery	The process of identifying, securing, reviewing information that is potentially relevant to the matter and producing information that can be utilized as evidence in the legal process.
Document Family	All parts of a group of documents that are connected to each other for purposes of communication; e.g. an email and its attachments.
Email Threading	The process of regrouping emails together that may have been found separately in collected data. Facilitates the ability to look through entire conversations for relevancy and privilege categorization.
e-Disclosure	The eDiscovery process as it is practiced in the European Union.

Electronic discovery	eDiscovery, e-discovery. The process of identifying, preserving, collection, preparing, reviewing and producing ESI in the context of a legal process.
Electronic evidence	Information that is stored in an electronic format used to prove or disprove the facts of a legal matter.
Email	An electronic communication sent or received via a data application (e.g. MS Outlook, Lotus Notes, Google Gmail).
ESI	Electronically Stored Information
Filtering	The process of specific parameters to remove documents that do not fit those parameters in order to reduce the volume of the data set (e.g. date ranges and keywords).
Forensics	A handling of ESI including collection, examination and analysis, in a manner that ensures its authenticity, so as to provide for its admission as evidence in a court of law.
FRCP	Federal Rules of Civil Procedure, the rules that govern eDiscovery and other aspects of the civil legal process.
Hash	An algorithm that generates a unique value for each document. It is often referred to as a digital fingerprint and is used to authenticate documents and identify duplicate documents.
Image (Drive)	Mirror image. To make an identical copy of a drive including its empty space.
Image (File)	To make a picture copy of a document. The most common image formats in eDiscovery are tiff and pdf.
Legacy Data	Data whose format has become obsolete, making it difficult to access or process.
Legal Hold	A communication requesting the preservation of information that is potentially relevant to current or a reasonably anticipated legal matter.
Load File	A file used to import data into an eDiscovery system. It defines document parameters for imaged documents and often contains metadata for all ESI it relates to.
Media	The device used to store electronic information, e.g. hard drives, backup tapes, DVDs.
Metadata	Often referred to as data about data, it is the information that describes the characteristics of ESI, e.g. sender, recipient, author, date. Much of the metadata is not accessible to non-technical users.
Native Format	A computer file that is maintained in the format in which it was created. This format preserves metadata and details about the data that might be lost when the documents are converted to an image format such as TIFF or PDF. Spreadsheets, for example, are much more valuable in their native format. Without the native file, you can't view formulas or pivot tables that might exist in a spreadsheet.
Near-duplicate	Two or more files that contain a specified percentage of similarity. Also, the process used to identify those nearly-identical files.
Normalization	Reformatting data so that it is stored in a standardized format.
OCR	Optical Character Recognition is the process of converting images of printed pages into electronic text.
Parent Document	A document to which other documents/files are attached.

Predictive Coding	A document categorization process that extrapolates the tagging decisions of an expert reviewer across a data set. It is an iterative process that increases accuracy with multiple training passes.
PST	PST is a Microsoft Outlook email store. Multiple .pst files may exist and contain archived email.
Precision	In search results analysis, precision is the measure of the level of relevance to the query in the results set documents.
Processing	The eDiscovery workflow which ingests data, extracts text and metadata, normalizes the data. Some systems include the data indexing and deduplication in their processing workflow.
Production	The delivery of documents and other data that meet the criteria of the discovery request.
Recall	In search results analysis, recall is the measure of the % of total number of relevant documents in the corpus returned in the results set.
Redact	To intentionally conceal, usually via an overlay, portions of a document considered privileged, proprietary or confidential.
Search	The process of looking within a data set using specific criteria (a query). There are several types of search ranging from simple keyword to concept searches that identify documents related to the query even when the query term is not present in the document.
Slack space	The unused portion of a disk that exists when the data does not completely fill the space allotted for it. This space can be examined for otherwise unavailable data.
Spoliation	The destruction or alteration of data that might be relevant to a legal matter.
Structured data	Data stored in a structured format such as a database. Structured data can create challenges in eDiscovery. <i>See Unstructured data</i>
TIFF	Tagged Image Format is a common graphic file format. The file extension related to this format is .tif or .tiff
Unallocated space	Most often, this is space created on a hard drive when a file is marked for deletion. This space is no longer allocated to a specific file. Until it is overwritten, it still contains the previous data and can be retrieved.
Unicode	The code standard that provides for uniform representation of character sets for all languages. It is also referred to as double-byte language.
Unstructured data	Data that is not stored in a structured format such as word processing documents and presentations.