# 52 eDiscovery Terms That You Should Know

## A

**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 repository for the storage of records and files

**Attachment**: A document or file that is connected to another document or file either externally, e.g. a document connected to an email, or embedded, e.g. an image in a word processing document.

**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.

## B

**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

## C

**Chain-of-Custody**: The processing, tracking and recording the movement, handling and location of electronic evidence chronologically from collection to production. It is used to verify the authenticity of the ESI.

**Child Document**: A file that is attached to another communication file. E.g. the attachment to an email or a spreadsheet imbedded in a word processing document

**Container File**: A single file containing multiple documents and/or files, usually in a compressed format; e.g. zip, rar, pst

**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.

## D

**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 documents generated by a computer system from those created by a user. This automated process utilized a list of file extensions developed by the National Institute of Standards and Technology.

**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

## E

**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 or investigative process

**Electronic evidence**: Information that is stored in an electronic format. This is used to prove or disprove the facts of a legal matter.

**Email**: An electronic communication sent or received via a data application designed for that purpose (e.g. MS Outlook, Lotus Notes, Google Gmail)

**ESI**: Electronically stored information

## F

**Filtering**: The process of applying specific parameters to remove groups of documents that do not fit those parameters, in order to reduce the volume of the data set, e.g. date ranges and keywords

**Forensics**: The 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.

## H

**Hash**: An algorithm that generates a unique value for each document. It is referred to as a digital fingerprint and is used to authenticate documents and to identify duplicate documents.

## I

**Image (Drive)**: To make an identical copy of a drive including its empty space. "mirror image"

**Image (File)**: To make a picture copy of a document. The most common image formats in ediscovery are tiff and pdf.

## L

**Legacy Data**: Data whose format has become obsolete.

**Legal Hold**: A communication requesting the preservation of information that is potentially relevant to current or a reasonably anticipated legal matter and the resulting preservation.

**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.

## M

**Media**: The device used to store electronic information, e.g. hard drives, back up 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.

## N

**Native Format**: A 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 image format, e.g. pivot tables in spreadsheets.

**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

## O

**OCR**: Optical Character Recognition is the process of converting images of printed pages into electronic text

## P

**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.

**Precision**: In search results analysis, precision is the measure of the level of relevance to the query in the results set of documents. *See Recall*

**Processing**: The ediscovery workflow which ingests data, extracts text and metadata, and normalizes the data. Some systems include the data indexing and de-duplication in their processing workflow.

**Production**: The delivery, to the requesting party, of documents and ESI that meet the criteria of the discovery request

## R

**Recall**: In search results analysis, recall is the measure of the percentage of total number of relevant documents in the corpus returned in the results set. *See Precision*

**Redact**: To intentionally conceal, usually via an overlay, portions of a document considered privileged, proprietary or confidential

## S

**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. *See Unstructured data*

## T

**TIFF**: Tagged Image Format is a common graphic file format. The file extension related to this format is .tif.

## U

**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 often 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

Complete Discovery Source (CDS Legal) is a leading provider of eDiscovery services and a trusted advisor to more than 75% of the Am Law 100 and a quarter of the Fortune 500. CDS uses the most advanced, tested, and defensible technology to support all stages of the EDRM including planning, early case assessment, processing and production, review hosting, managed review, and cross-border global coverage.

## Review Hosting and Managed Review

If you're working around the clock, so are we. We support our highly secure SAS 70 Type II data centers 24/7. And rest assured that we employ the most advanced technologies and protocols to protect sensitive data from damage by malice, accident or error.

CDS partners with the industry's leading review tool developers and contract review staff to meet your case's specific needs. For complex projects, our experienced PMs and Director of eDiscovery Strategy work with case teams to tailor workflows—from reviewer training to trial preparation.

CDS' National Data Center is SAS 70 Type II audited. Our three locations are available 24/7.

## Global Coverage and Trans-border eDiscovery/eDisclosure

The world is getting smaller, but the demands of global electronic discovery are growing. Multi-national corporations involved in U.S. litigation can be required to collect data that is subject to EU directives. CDS' international eDisclosure program merges EU-based eDisclosure specialists with our proven U.S. based eDiscovery firm, to seamlessly carry you through the process. And CDS' Safe Harbor certification ensures that data which reaches the U.S. remains compliant with EU privacy regulations. From collection to production, our solutions are ready for your overseas needs.

Our Unicode-compliant technology provides multi-lingual capabilities. It automatically identifies the languages within documents and provides detailed reports. Reviewers can search documents in Roman, Asian, and Cyrillic languages.

---

*"CDS has repeatedly proven its reputation as an industry leader, a partner guaranteed to tackle and meet any project and deadline in a cost effective manner. [Our] entire project had to be completed within two days time . . . and they completed the project early . . . CDS is a highly valuable, respected, and competitive asset in the e-discovery community."*

**— E-DISCOVERY PROJECT MANAGER, TOP 50 LAW FIRM**

*"CDS has been a tremendous help and life saver . . . we have used them to migrate several databases from other platforms because of their stellar support and services . . . they have lived up to what they promised."*

**— LIT SUPPORT MANAGER, AM LAW 100 FIRM**

---

Relativity s3).3) s))3

New York Law Journal Rankings — Top Three Best eDiscovery Provider — Complete Discovery Source

GSA Schedule Contract 35F-0480R

We self-certify compliance with U.S. • EU SAFEHARBOR U.S. DEPARTMENT OF COMMERCE

CDS COMPLETE DISCOVERY SOURCE

**www.cdslegal.com**
**(855) 813-0700**

CDS Headquarters
345 Park Avenue
New York, NY 10154
(212) 813-7000

CDS Washington
1050 Connecticut Avenue NW
Washington, DC 20036
(202) 822-5999

CDS Chicago
70 West Madison St.
Three First National Plaza
Suite 1400
Chicago, IL 60602-4270
(312) 756-7301

CDS San Francisco
425 Market Street
Suite 2200
San Francisco, CA 94105
(415) 692-5595