

Glossary of e-Discovery Terms

By Jamie B. Schwartz

For the practitioner who is unfamiliar with the emerging e-discovery terms, there are many resources available. The following are a few key e-discovery terms adopted from the *Sedona Conference Glossary: E-Discovery and Digital Information Management*, May 2005.

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 restoration or reconstruction.

Backup Data.

Backup data is an exact copy of system data that serves as a source for recovery in the event of a system problem or disaster. Backup data is generally stored separately from active data on portable media.

Data Filtering.

Data filtering is the process of identifying for extraction specific data based on specified parameters (e.g., by key word, file type, or name).

De-Duplication.

De-duplication is the process of comparing electronic records based on their characteristics and removing or marking duplicate records within the data set.

Legacy Data, Legacy System.

Legacy data is information that an organization may have invested significant resources developing, and while it has retained its importance, the information has been created or stored by the use of software and/or hardware that has become obsolete or replaced. Legacy data may be costly to restore or reconstruct when required for investigation or litigation analysis or discovery.



Metadata.

Metadata is information about a particular data set or document that describes how, when, and by whom it was collected, created, accessed, modified, and how it is formatted. It can be altered intentionally or inadvertently. It may also be extracted when native files are converted to image. Some metadata, such as file dates and sizes, can be seen easily 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.

Native Format.

Electronic documents have an associated file structure defined by the original creating application. This file structure is the document's native format.

PDF (Portable Document Format).

A PDF captures formatting information from a variety of applications in such a way that they can be viewed and printed as they were intended in their original application by practically any computer, on multiple platforms, regardless of the specific application in which the original was created. PDF files may be text-searchable or image-only.

PST.

PST is a Microsoft Outlook email store. Multiple .pst files may exist and contain archived email.

Sampling.

Sampling usually refers to the process of testing a database for the existence or frequency of relevant information. It can be a useful technique in addressing a number of issues relating to litigation, including decisions about what repositories of data are appropriate to search in a particular litigation and determinations of the validity and effectiveness of searches or other data extraction procedures.

TIFF (Tagged Image File Format).

A TIFF file has a .tif extension. Images are stored in tagged fields, and programs use the tags to accept or ignore fields, depending on the application.

JAMIE B. SCHWARTZ is an electronic discovery specialist with LexisNexis Applied Discovery.

LexisNexis is a Sponsor of the Section of Litigation, and this article appears in connection with the Section's sponsorship agreement with LexisNexis. Neither the ABA nor ABA Sections endorse non-ABA products or services.