Over the next several weeks the *E-Discovery Alert* will focus on the strategy and tactics for handling sixteen specific ESI issues throughout pretrial discovery. Whether it is a "meet and confer" or request for production these are the critical issues to focus in requesting or producing ESI. The legal issue excerpts will be derived from the
Best Practices Guide for ESI Pretrial Discovery - Strategy and Tactics
(2008-2009). The
Guide
is cross-referenced and hyperlinked with the
Arkfeld on Electronic Discovery and Evidence
(2nd ed.) treatise and part of the CD-ROM.

ISSUE: IS IT CRITICAL TO UNDERSTAND THE TERMINOLOGY AND DIFFERENT CLASSIFICATIONS OF ESI?

ANSWER: YES

Resources:

- See "Forms or Forms of ESI." Click on the link for a set proposed terms to clarify the questions raised by the "form or forms" reference under Rule 26(f) and Rule 34 of the FRCP. These terms can be freely used by attorneys and other legal professionals as a "common glossary" to discuss the "form" of disclosure of ESI.

- Excerpt from *Best Practices Guide for ESI Pretrial Discovery - Strategy and Tactics* (2008-2009)
, § 3.3,
Terminology and Classification of ESI

§ 3.3 TERMINOLOGY AND CLASSIFICATION OF ESI

ESI is classified according to its type, intended use or as to the means by which such information can be accessed. Understanding the classifications of electronic information is important in mastering electronic discovery principles.

Unfortunately, there is not a consensus as to the definition of different e discovery terms. Confusion occurs when parties use terms such as "metadata," "load files," "active files," and many others. It is strongly suggested that parties in litigation reach agreement, as soon as practicable, as to the terminology they employ to guard against disputes over the discovery or production of ESI.

A. Legal Terminology and Classification

How information is classified may have a determinative effect on how the courts decide a cost-shifting request. For example, the court in the seminal decision *Zubulake v. UBS Warburg LLC,* 217 F.R.D. 309, 319-320 (S.D.N.Y. 2003) identified five categories of electronic information and classified each as to the accessibility of the information.

Whether electronic data is accessible or inaccessible turns largely on the media on which it is stored. Five categories of data, listed in order from most accessible to least accessible, are described in the literature on electronic data storage:

1. Active, online data: "On-line storage is generally provided by magnetic disk. It is used in the very active stages of an electronic records [sic] life - when it is being created or received and processed, as well as when the access frequency is high and the required speed of access is very fast, i.e., milliseconds." Examples of online data include hard drives.

2. Near-line data: "This typically consists of a robotic storage device (robotic library) that houses removable media, uses robotic arms to access the media, and uses multiple read/write devices to store and retrieve records. Access speeds can range from as low as milliseconds if the media is already in a read device, up to 10-30 seconds for optical disk technology, and between 20-120 seconds for sequentially searched media, such as magnetic tape." Examples include optical disks.

3. Offline storage/archives: "This is removable optical disk or magnetic tape media, which can be labeled and stored in a shelf or rack. Off-line storage of electronic records is traditionally used for making disaster copies of records and also for records considered 'archival' in that their likelihood of retrieval is minimal. Accessibility to off-line media involves manual intervention and is much slower than on-line or near-line storage. Access speed may be minutes, hours, or even

days, depending on the access-effectiveness of the storage facility." The principled difference between nearline data and offline data is that offline data lacks "the coordinated control of an intelligent disk subsystem," and is, in the lingo, JBOD ("Just a Bunch of Disks").

4. Backup tapes: "A device, like a tape recorder, that reads data from and writes it onto a tape. Tape drives have data capacities of anywhere from a few hundred kilobytes to several gigabytes. Their transfer speeds also vary considerably . . . The disadvantage of tape drives is that they are sequential-access devices, which means that to read any particular block of data, you need to read all the preceding blocks." As a result, "[t]he data on a backup tape are not organized for retrieval of individual documents or files [because] . . . the organization of the data mirrors the computer's structure, not the human records management structure." Backup tapes also typically employ some sort of data compression, permitting more data to be stored on each tape, but also making restoration more time-consuming and expensive, especially given the lack of uniform standard governing data compression.
5. Erased, fragmented or damaged data: "When a file is first created and saved, it is laid down on the [storage media] in contiguous clusters . . . As files are erased, their clusters are made available again as free space. Eventually, some newly created files become larger than the remaining contiguous free space. These files are then broken up and randomly placed throughout the disk." Such broken-up files are said to be "fragmented," and along with damaged and erased data can only be accessed after significant processing.

Of these, the first three categories are typically identified as accessible, and the latter two as inaccessible. The difference between the two classes is easy to appreciate. Information deemed "accessible" is stored in a readily usable format. Although the time it takes to actually access the data ranges from milliseconds to days, the data does not need to be restored or otherwise manipulated to be usable. "Inaccessible" data, on the other hand, is not readily usable. Backup tapes must be restored using a process similar to that previously described, fragmented data must be de-fragmented, and erased data must be reconstructed, all before the data is usable. That makes such data inaccessible.
When conversing with a computer forensic specialist or others, the following classifications are generally used. Grace V. Bacon, *The Fundamentals of Electronic Discovery,* 47 Boston Law Journal 18 (2003).

A. DATA FILES
Data files are the basic information that computer systems store. There are four general types of data files: (1) active data; (2) replicant data; (3) backup data; and (4) residual data.

"Active data" is the information currently accessible on a computer, such as word processing documents, spreadsheets, databases, e-mail messages and electronic calendars. Generally, active data is relatively simple to access through the use of a computer's file manager program. It can be found on an individual's office desktop computer, laptop, home computer, an assistant's computer, a PDA and the network file server. Moreover, because users frequently create special files or folders in which to store e-mails or other electronic documents that pertain to a particular subject matter, active data will usually be fairly easy to sort for relevant information. Most computer programs also contain search engines that can be used to narrow

the scope of potentially relevant documents.

2. Replicant Data
"Replicant data" (or "archival data") is the information a computer automatically backs up as you work on a file. These backed up files are created and saved in order to recover data that may be lost due to a malfunction or power loss. Replicant data is useful because it creates a copy or several copies of a document that the user may not erase. In fact, the user may not even be aware of these "file clones" because they are generally stored in a different directory than active data. On most networked systems, this replicant data is stored on the hard drive as opposed to a centralized network file server. Consequently, a document, or part of it that was purged from a server, may be retrievable from a user's hard drive.

3. Backup Data
"Backup data" is information copied to a removable medium in the event of a system failure. Most businesses have their networks backed up on a routine schedule, while individual users may or may not backup their information. Thus, one can find backup data on system-wide backup tapes, recovery backup tapes that may be stored off site, and on personal backups such as computer disks.
Backup data is particularly useful in that it provides historical snapshots of the data stored on a system on the specific day the backup was made, allowing one to obtain information regarding the progress of a matter. On the flip side, because backup tapes contain a large amount of data, it is frequently time consuming and expensive to restore this data in order to review the material pertinent to your case.

4. Residual Data
Unlike general "paper" discovery, electronic documents thought to be lost or destroyed are more often than not recoverable, yielding what is often an untapped source of information in a case. Simply pressing the "delete" button does not mean that the document is no longer on the computer. "Residual data" is information that is actually recoverable even though an attempt has been made to "delete" the document. When a file is "deleted," the computer makes the space occupied by that file available for new data. Unless that space is "re-written," the so-called deleted document is generally recoverable by using "undelete" or "restore" commands contained in some systems' operating software or through other programs."

* * * * * * denotes content that has been omitted