What is an Active File Collection?

Description

Active File Collection refers to the collection of files that are active (not deleted) and pertain to a legal matter or legal hold. In most civil litigation cases, extensive forensic investigations that look at deleted files are unnecessary or too expensive. Thus, most ESI collections are active file collections and/or email collections.

How active file collections are performed

Active files are those that can be seen by normal users. They may include hidden or system files, but they do not include the computerâ€™s Random Access Memory or any deleted files. Files in the Windows Recycle Bin are considered active files and are subject to collection using active file collection methods.

The first step is defining which files need to be collected. This definition can range from â€œeverythingâ€• to files of a few specific types containing only certain key words. Since the cost of processing is usually related to the size of the data being processed, it is generally more economical to be as specific as possible without leaving out relevant files.

Once the files have been identified, it is mostly a matter of copying them in a manner that both avoids spoliation and provides a means of certifying the contents of the copies.

What to remember

The one thing to remember about active file collections is that they can be a potential minefield of spoliation. To avoid this, use software that is designed to preserve the metadata, the timestamps, and the data within the copied files. Some products, such as SafeCopy 2 from Pinpoint Labs are designed specifically for this purpose. Others, like Harvester, also from Pinpoint Labs, offer this feature as well as the ability to cull data by key word search and also support deduplication, email, and deNISTing.

The most important aspects of active file collections are preservation and validation.

Preservation refers to the preservation of the file data, its timestamps (when the file was created, last modified, and last accessed), and any other metadata contained within the file. If any of this data is compromised, the usefulness and admissibility of the file comes into question.

Validation refers to the ability to certify that the contents of the copy are the same as the contents of the original. This is usually done using a hash (analogous to a fingerprint of the files data). It may also be done using a bitwise comparison of the data in both the file and the copy, but since this method requires the same amount of storage as the files themselves and offers no means of independent verification, it is not in common use.

Date

11/01/2025
Date Created
11/30/2010