PINPOINT LABS

PRESERVE · COLLECT · DISCOVER

A NO-NONSENSE GUIDE TO E-DISCOVERY COLLECTIONS

Learn a Few Simple Steps That Prevent Spoliation,
Incomplete Productions and Increase Potential Evidence

Version 2.0 dated 11/01/2015

Jon Rowe, CEO
Computer Forensic Examiner
Pinpoint Labs
jon.rowe@pinpointlabs.com
www.pinpointlabs.com

IMPORTANCE OF ELECTRONIC FILE COLLECTIONS.

Incomplete and undocumented electronic discovery collections occur every day and the results are costly to both clients and their legal counsel. In typical litigation, clients expect their attorneys to guide them through the entire process. After all, litigation is what attorneys do for a living. However, litigation is a burden to clients in that it is both costly and distracts client staff from the client's core business. In the past, few would argue that the failure to properly advise a client on how a file cabinet of physical documents should be handled in a litigation was potential malpractice. Now that most documents are electronic and the file cabinets are computer networks, the failure to properly advise a client on handling electronic data is potential malpractice in the form of spoliation and missing evidence.

Advising clients about proper file collection methods must be considered from the beginning of each case. Using processes and applications that preserve and verify collected electronic files with minimal impact on client systems is critical. Without them, creating defensible and verifiable electronic discovery productions and evidence authentication is difficult, if not impossible.

COMMON PROBLEMS IN FILE COLLECTIONS.

Normally, a discovery request is created and the corporate IT department or other client employees copy relevant files or directories to disks or a USB drive. From the client's perspective, this is the least costly way to collect data. However, it can be the most costly way in the long run. The integrity of files collected from corporate servers and client machines are in jeopardy because many electronic document collections are completed using tools that lack the ability to confirm results and properly document the process.

Attorneys need to advise clients as to proper collection techniques. While not every case warrants a full third party forensic collection, every case does warrant a defensible and verifiable electronic data collection process. Discussion about when a full third party forensic collection is needed is beyond the scope of this article but the possible need must be considered before the decision is made to move forward with a different class of collection.

INCOMPLETE FILE COLLECTIONS

Electronic file collection projects can take many hours or days and contain hundreds of thousands, if not millions, of files. The software used to copy and burn files often lack a verification process; therefore, files that are skipped, partially copied or corrupted go unnoticed.

Incomplete and corrupted file collections pose an unseen danger as reviewers may never know that a relevant file was unavailable or unsearchable. The best way to ensure that all relevant files are identified, properly copied and delivered without error includes:

- Hash verification for every file.
- Log incomplete copies, files in use or skipped files.
- Maintain descriptive error logs.
- Proactive error reporting and feedback.
- Verification (chain of custody) log.

Recommending that clients use file collection methods with these options helps ensure all electronically produced files are intact and available for review.

7 COMMON PROBLEMS IN E-DISCOVERY COLLECTIONS

1) FILES IN LONG PATHS ARE SKIPPED

Anyone who has worked in e-discovery for any length of time has encountered problems working with files that are located in paths greater than 255 characters. Microsoft Windows and many other applications can't access files where the total number of characters (including folder and file name) exceeds 255.

It is common for a custodian's computer or the company file shares to store files in long paths. Missing all files that are stored in long paths is a frequent problem when backing up or collecting files. Often there is no warning or notification that a long file path was encountered; therefore, users are unaware that potentially critical information was not captured.

The problem is compounded when opposing counsel begins asking questions as to why specific documents weren't produced. One of the hardest obstacles to overcome is the perception that documents were purposely withheld or a less than credible collection process was used to produce electronic data.

2) ALTERED FILE AND FOLDER TIMESTAMPS

During normal file copy operations, Microsoft Windows creates new and updates existing file system timestamps on the new copies as well as the originals. This often causes problems later when the file metadata is imported into attorney review platforms and needs to be organized or searched.

Creating an accurate timeline of events is a critical component in the discovery process and not having access to the original file timestamps can quickly become an issue. Microsoft Office and other applications files have internal metadata that can work as a backup to help determine the date a file was created or edited. However, internal timestamps will still differ from the file system timestamps.

Additionally, many file types do not contain internal metadata and the only record of when a file was created is contained within the file system metadata. It is critical to use copy utilities that can preserve both file and folder timestamps during an e-discovery collection to ensure proper timelines can be created during review.

3) INCOMPLETE COLLECTION PROJECTS

Recovering from interruptions, identifying missed files and easily correcting is not possible using Microsoft Windows file copying as well as many other free and paid copy applications. Additionally, several disk imaging applications must restart if an error occurs while writing files to the container.

Network outages, computer restarts and end user job cancellation are very common during e-discovery collections. The ability to easily identify which files had trouble as well as which ones remain to be copied is critical to ensure the collection is defensible. Those who have been involved in any kind of large scale collection or backup project are intimately aware of these issues and most likely spent many hours or days attempting to complete a project and ensure all files were copied. Yet they are often not 100% confident that the job completed successfully and has detailed logs to confirm the process.

4) CLIENT SYSTEM MODIFICATION

Current collection efforts should include using applications and processes that preserve the native states on servers and individual client PC's. However, many file collection utilities require an installation on corporate systems, which can introduce multiple obstacles:

- Corporate IT policies may prevent installing new software without prior approval.
- Transferring software licenses between systems may require purchasing multiple copies or contacting the developer.
- Collection software may not have all file dependencies on the target computer.

File collection software should ideally be portable and run without installation. The ***"zero footprint"*** software option preserves the native state of the servers and client machines.

5) OVER-COLLECTING

Due to the amount of data that can now reside on corporate servers and individual custodian PC's, many companies and legal counsel seek ways to reduce the amount of data at the point of collection. In many cases, keyword search, date ranges, deduplication, deNisting and other criteria are applied. It is not uncommon to have the data produced reduced by 90% and more, which can be a considerable cost savings.

Culling data at the point of collection can make a lot of sense. However, before deciding to apply keywords at the point of collection, there are certain dangers that need to be considered. The complete keyword list needs to be finalized if there is only one chance to collect information from the producing party. Why is settling on keywords so critical? The answer is that any additional keyword revisions will only apply to the collected information and not across all original sources.

To further explain, assume a corporation has 10 terabytes of information that needs to be collected and only 500 gigabytes is produced that matched keyword searching and other criteria. The 500GB is what will be available for searching during e-discovery processing and review. If additional keywords are considered relevant and need to be applied to the collected data, there will be 9.5 terabytes of data that is now excluded from the search because it is still sitting back on the corporate servers and custodian PC's because it never matched the original keywords

and wasn't produced. Therefore, it is critical that the attorneys understand that additional changes will only be applied to the already filtered information and not the entire universe of the original documents.

6) FILES SKIPPED DURING KEYWORD SEARCHING

Leaving potentially relevant documents behind due to encryption is another issue that occurs when keyword searching at the point of collection. Why? Corporate IT departments and IT legal professionals often use programs to search their networks and PC's for files containing keywords and then produce a list of files that need to be collected. However, they don't realize that any files that were password protected or encrypted were skipped because a keyword search could not be applied.

To further explain, when a user applies a password to a file, the contents are "scrambled," for lack of a better term, which is what allows the file to be protected so the contents are easily viewed by other applications. When the keyword search process encounters the file, it will not be able to see the original content or find matching data.

Therefore, any potentially relevant files that are encrypted will be left behind unless the application searching the data is designed to identify these encrypted files and ensure the user knows they exist. Why not just decrypt and search them on-the-fly to determine if they are a keyword match? Great question! Decrypting a file could take anywhere from seconds to many years depending on the complexity. Clarifying if a file should be decrypted often needs to be discussed between a client and its legal counsel.

The most efficient way to handle encrypted files during collection is to identify, copy and create a list so they can be reviewed and determine if a decryption process needs to be used. Leaving them behind is not the answer; however, many processes in place do exactly that. If your company or legal counsel decides on a targeted collection, it would be advised to ask what process and applications are going to be used.

7) OPTIMAL KEYWORDS ARE MISSED

Keyword hit preview and reporting are very useful because they can list the files that are a match and provide a preview of what would be collected before copying. Often keywords are applied only to find out that the collected files and emails do not match expectations. When this happens, the keyword search criteria need to be altered and in some cases, going back to the sources is not an option.

INCONSISTENT RESULTS

When attorneys and their support staff are not involved in recommendations or implementing best practices for file collections (or fail to even know what the best practices are), the quality of the file productions can suffer and client claims for malpractice can result. When individuals responsible for file collections are not familiar with adequate collection tools, they may resort to file copy utilities that do not include verification or they do not know how to set the options.

Common copy utilities have dozens of options, which if not used in the right combination, can cause a number of errors. Additionally, there can be a higher likelihood of errors if multiple parties attempt to replicate the same settings.

It is important to ensure that file collections are consistent across multiple projects. Using intuitive tools that require minimal end user interaction is preferred.

AVOID COMMON COLLECTION PROBLEMS

As a result of the crippling issues identified in this article, a new breed of collection software was developed. You can learn more about **SafeCopy, Harvester, SharePoint Collector** and **Pinpoint Labs** at www.pinpointlabs.com. It focuses on forensically sound tools related to preservation, collection and filtering. Pinpoint Labs applications are intuitive, affordable and address common litigation support needs.

PINPOINT LABS SOFTWARE

SafeCopy has been relied on for almost a decade by legal IT, corporate security and computer forensic experts who need to confidently complete e-discovery productions. **SafeCopy Portable** and Server licenses allow users to easily and defensibly collect and back-up client data. http://pinpointlabs.com/sc2.html/

Harvester Portable enables users to filter and defensibly collect e-discovery files from a laptop, desktop or network location. It can be run from an external hard drive or a host computer. In addition to collecting data, **Harvester Portable** is also a very powerful culling tool that many use after collection to filter Microsoft Outlook PST's and loose files by keyword, date range and other criteria.
http://pinpointlabs.com/occh.html/

Harvester Server can be used to monitor jobs from a central location and remotely launch pre-configured collection job profiles on remote computers. In addition to collecting data, **Harvester Server** is also a very powerful culling tool that many use after collection to filter Microsoft Outlook PST's and loose files by keyword, date range and other criteria. Users are able to view real-time feedback as collection jobs start and to monitor their progress. Progress results can be emailed directly from Harvester Server as a Microsoft Excel spreadsheet, PDF, RTF and in many other formats. http://pinpointlabs.com/occh.html/

SharePoint Collector can quickly export document libraries, as well as perform targeted collections when needed, dramatically reducing the collection size and overall project costs. Common electronically stored information (ESI) relevant to a litigation is often stored in SharePoint sites, but companies struggle in meeting their production requests. They find the built-in search and retrieval features don't provide a means to satisfy their requirements and don't know where to turn.
http://pinpointlabs.com/sharepoint-collection.html/

SUMMARY

Many recognize that preserving, verifying and documenting electronic discovery collections confirms that relevant files are acquired. It also helps legal departments avoid spoliation and demonstrates to their clients they are implementing best practices. As inside counsel, general counsel and corporate IT departments learn more about litigation readiness, it becomes more important that their partnering legal departments keep abreast of the changes and are the ones leading the way.

Being proactive and recommending the proper methods and tools for ESI collections will ensure consistent results and provide a "heads-up" on any issues encountered. Many legal departments and service providers rely on Pinpoint Labs software tools for active file collections because collections results are confirmed, incomplete jobs are immediately reported and the process is thoroughly documented.

PINPOINT LABS

PRESERVE · COLLECT · DISCOVER

ABOUT PINPOINT LABS

Pinpoint Labs was founded by Jon Rowe and James Beasley, who are Computer Forensic Examiners. Their experience includes 20 years of litigation support and more than two decades in software development.

WWW.PINPOINTLABS.COM

▸ 8246 Oddo Circle ▸ Murray, NE 68409 ▸ 402.235.2381 ▸ 888.304.1096