

Is Your Data Collection Process

Legally Defensible?

By Rob Coviello*

It is never easy to forecast the path of litigation and determine the level of scrutiny an opposing party will use in examining a collection of electronically stored information. Regardless of the particular situation, one important factor to keep in mind is that electronic documents have unique metadata characteristics and their integrity can be compromised if handled improperly during the data collection phase. Because data collection occurs so early in the overall electronic discovery process, any mistake or oversight that occurs during this phase can adversely affect each subsequent phase of the electronic discovery process.

Why Maintain a Defensible Data Collection Plan?

The importance of maintaining a defensible data collection process was first recognized long before standard processes and commercial tools were available to preserve, examine and process electronic evidence. In the early days of computer forensics and electronic discovery, even before the existence of those terms, digital evidence was rarely contested. This is largely because there was limited

knowledge and understanding of it outside of the law enforcement community. Over time, knowledge and awareness of digital evidence began to increase. With that, data collection and forensic processes were being challenged by tech-savvy defense teams who knew just enough about electronic data to cast a reasonable doubt as to its integrity. The law enforcement community responded by refining

their processes to create repeatable, tested and defensible data collection and preservation protocols. These protocols would become generally accepted by the scientific community and would also endure trial testing over time. Specific defensible processes included the protocols used for the handling of digital evidence, maintaining chain-of-custody for digital evidence and the specific software and hardware tools used to collect and preserve digital evidence.

In a criminal case, where evidence is most likely to undergo intense scrutiny, the need for a defensible data collection process is obvious - one missed step in the collection process and an entire set of data may be disallowed as evidence. In civil litigation, however, the accepted data handling standard can sometimes be inadvertently held much lower by corporate legal departments and law firms. In many cases the importance of a defensible data collection process may not yet be completely understood. A full forensic bit-stream image of a custodian hard drive may not be deemed necessary to collect specific files, for example, but this is not an indication that the process used should not be legally defensible.

In order to recognize the ways in which a poor data collection process can adversely impact litigation, we must first look at the electronic discovery process and the importance of metadata information.

The Electronic Discovery Process— a Brief Summary

When faced with a discovery request, one of the initial actions in the electronic discovery process is to quantify and collect potentially relevant electronically stored information (ESI). Once this stage is complete, the next phase of the process involves the inventorying and unwrapping of the many logical layers of the ESI. One such example is extracting email messages from an electronic mailbox, then extracting the attachments of those email messages and so on. If a compromise has been reached with the requesting party allowing the use of filtering, such as by keyword or date, this step would follow. Only those documents and contextual relatives deemed relevant to the filtering criteria would then move forward to the next phase in the process and be prepared for access in a document review system.

The Importance of Metadata

Metadata consist of descriptive file information including creation and modification dates, file paths, author information and a variety of other data descriptors.

In most document review platforms, metadata are associated directly to its corresponding electronic document. Metadata can sometimes





If there is a chance that relevant files may have been deleted, it is important to move quickly while being diligent.

be relied upon to determine a document's category or even possibly its relevance. Some regulatory agencies have even begun to request that certain metadata information be included with electronic document productions in the form of a load file, a more advanced approach than the older standard of printing to paper or TIFF-only deliverables. The importance of an electronic document's metadata continues to become more understood in the civil litigation world. Just as the context of a document or email can be used to establish who knew what, metadata can often be used to establish *when* they knew it.

Metadata can be critical when establishing a timeline of events, and it is this very same information that can be unintentionally altered if data collection is not performed properly. For this reason, there are a few key points to keep in mind when initiating the data collection process in order to maintain defensibility.

Clicks and keystrokes can destroy data integrity. Merely clicking on and opening a file on a subject's computer can change metadata information about a file, such as the file's last access time, that previously could have been recoverable. It would only take one or two more clicks or keystrokes to alter that same file's last modification time, not to mention stored information about who was the last person to save that file. An opposing party would relish the opportunity to attempt to discredit the integrity of an entire electronic data set if they were to discover this type of metadata alteration resulting from a seemingly harmless perusal of files.

Copy-and-paste does not copy everything. Although a progress bar may report that files are copying over to a CD, and those files haven't been opened at all, there is still a potential for data loss. Even when data are being copied to media that may ultimately become read-only and not allow future modification, the process used to create that copy is the most critical component. Simply put, copying and pasting from a subject's computer to destination media, such as a CD, will not copy over some critical metadata components. Although a file's hash value—also known as a digital fingerprint—may verify after a simple copy and paste, the creation date, original file path and folder information, among other items, have now been lost and left behind.

Industry-accepted tools and processes have been accepted for a reason. There are a variety of ways to collect data and maintain their integrity. At the highest level, creating a full bit-stream image of the source media while using appropriate write-block devices and procedures will capture the most information. This includes deleted files and unallocated disk space where fragments of once-active files may reside. There are a number of proven tools available on the commercial market that can create full bit-stream images. In situations where full bit-stream images may not be necessary, there are also a number of tools available to selectively capture active files while retaining metadata information. The use of these tools can still be legally defensible, but the important thing to remember is that the tool and process should be proven and scientifically tested in order to defend its validity. The process also needs to be appropriately documented. For this reason it is often best to utilize a professional with the appropriate training and knowledge of data collection processes. An internal IT resource may be well versed in technology but not appropriately qualified or certified to handle data collection as part of a legally defensible process, and testify if required to do so.

Move quickly. If there is a chance that relevant files may have been deleted, it is important to move quickly while being diligent. In these instances, a forensically sound bit-stream image of a computer's hard drive or other media may be the only type of collection method that will allow the possible recovery of deleted files. When a file is normally deleted, it resides on the hard drive until another piece of data writes over it, so the chance of recovering deleted data significantly lowers over time. There could also be valuable metadata associated with deleted items that could be lost if the once-recoverable file is overwritten.

Maintain chain-of-custody. When handling a subject's computer, it is extremely important to document its possession, control and transfer. Even if the data collection is not expected to become part of a criminal litigation, for which the burden of proof may be much higher, it is important to document who did what, when, where and how it was done. In today's corporate world, where employees rely heavily on their laptops and computers, it may not be practical or cost effective for a corporation to seize and store every hard drive from which data need to be collected. Often times it is a verified fo-

rensic image of the hard drive that is collected and stored as best evidence. Chain-of-custody documentation should clearly define what the piece of evidence is, when it was acquired, who acquired it and how it was acquired. Information about the original computer that the digital evidence was collected from should also be recorded, such as system date and time, serial number, model number and how it was received. Chain-of-custody is often the key link between a piece of data and the person whose computer it resided on. This custodian information is often associated to documents in a document review system as additional metadata.

Alternative Data Sources. Depending on the scope of a data collection request, there may also be a need to acquire data that resides on a network file server, email system, or even backup tapes. Many of the same challenges with maintaining data and metadata integrity apply to data collected from these locations, and the same precautions must be considered. An accepted preservation method should be used to verify the integrity of the collected data. Chain-of-custody should always be maintained and there should be appropriate documentation of the process used.

Considerations for Self-Collection

Some corporate legal departments and law firms are exploring the idea of custodian self-collection or self-preservation. In these instances a custodian would be asked to prepare a CD, DVD, or other media that contains documents that they have deemed relevant to a particular document request. The copy-and-paste or drag-and-drop method is often used in these scenarios. While cost efficient, this certainly is not a forensically sound or legally defensible method of data collection. Critical metadata items can be altered when performing this type of data collection. In addition, any alteration of files or file properties may be a potential cause for future spoliation claims.

When document collection consisted mainly of paper documents, there may have been no harm seen in a custodian placing documents into a box after they had received a document request, and then sending the box off to counsel. Some would argue, however, even that approach relied too heavily on the good-faith effort of a custodian. But at least in the example of a paper document col-

lection there was little-to-no risk of the unintentional alteration of a document or its properties. Electronic documents do not share this fail-safe characteristic. Even a good-faith collection effort on the part of a custodian can alter an electronic document's metadata and other information.

Conclusion

The method by which data are collected will affect each subsequent stage of the electronic discovery process. In a sense, data collection can be considered a foundation for the electronic discovery stages that follow. If done correctly, there is a solid base from which to proceed; if not, there is an inherent downstream potential for disaster. Date culling and filtering can be impacted if metadata were inadvertently altered during an improper collection. Subsequently, a document review team may be reviewing documents that should not have made it through the filtering process, or missing documents that never made it through because of metadata alteration during the collection phase. Files that are being viewed in a document review platform may also be associated with incorrect metadata information if the metadata were altered during the collection phase. For these reasons, and to protect all parties involved in litigation, a proven and defensible method should always be used to collect data.

ROB COVIELLO is the Director, Discovery Solutions at TechLaw Solutions. Rob has coordinated and managed large-scale electronic discovery processing projects, including the design and implementation of standard operating and quality control procedures and establishing eDiscovery best practices. He has coordinated the forensic acquisition and analysis of digital evidence gathered from various domestic and international locations and directly led forensic data collection initiatives and electronic production requests in response to various state and federal government subpoenas. Rob is an EnCase-certified computer forensic examiner and a member of The Association of Certified Fraud Examiners.

