

PRESERVING ELECTRONICALLY STORED INFORMATION UNDER THE NEW RULES

by

Gregory L. Fordham, CPA, CIA, CCE, Sec+, MCP



www.knfcon.com
2550 Northwinds Parkway, Suite 275
Alpharetta, GA 30004
800-335-1188

PRESERVING ELECTRONICALLY STORED INFORMATION

PRESERVING ELECTRONICALLY STORED INFORMATION
UNDER THE NEW RULES

by

Gregory L. Fordham, CPA, CIA, CCE, Sec+, MCP

The preservation of Electronically Stored Information (ESI) is increasingly important. A party is required to preserve relevant evidence.¹ Relevant evidence is defined as “evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.”²

In the event a party fails in its duty, it can be subjected to sanctions.³ In recent times numerous decisions have been rendered where the spoliation of ESI has resulted in sanctions having devastating results to the offending party.

The need to preserve electronic data is greater than with paper documents, since electronic data is much more fragile than paper documents. Electronic data is more fragile than paper

documents because it is easier to change the content of electronic documents than the content of paper documents and because electronic data is more complex than paper documents. The evidentiary qualities of paper documents is limited to the content and media of the document itself. With respect to electronic documents, the evidentiary qualities extend beyond the document itself and include the filing system of the media being used to store the document as well as the electronic media on which the document is stored. For example, key date stamps indicating the documents last usage are stored in the filing system of the storage media and not necessarily in the document itself. The last accessed date stamp can be changed by a number of computerized processes simply by continuing to operate the computer. Similarly, earlier versions of the document could reside in unallocated portions of the storage media and would be available for examination until overwritten through routine use of the computer. Thus, in both cases discussed, the continued operation of a computer can destroy potential evidence.

There has been some debate concerning the breadth of preservation for ESI. Since under Rule 26(b)(2)(C) not all ESI is discoverable, the debate has centered on whether all ESI should be preserved. For example, in illustration I of Principle 5(a), Scope of Preservation Obligation, the Sedona Principles state,

“Absent awareness of a reasonable likelihood that specific unique and relevant information is contained only on a backup tape, there is no violation of preservation obligations, because the corporation has an appropriate policy in place and the backup tapes are reasonably considered to be redundant of the data

saved by other means.”

While there may be situations where the backup tape data is redundant and where preservation of the data is unnecessary that is not generally the interest in backup tapes. Indeed, the interest in backup tapes is that they contain data that is not currently available from other sources and thus, not redundant.

Similarly in Principle 9(b), Deleted Data and Residual Data, the Sedona Principles state,

“Absent specific circumstances, organizations should not have to preserve deleted or residual data. While most computer systems will have a plethora of data that could be “mined”, there should not be routine authorization for such forensic recovery. If, as usual, deleted and residual data are not accessed by employees in the ordinary course of business, there is no reason to require the routine preservation of such data. The relevance of the data will be marginal at best in most cases, while the burdens involved will usually be great. In exceptional cases, however, there may be good cause for targeted preservation of deleted and residual data.”

Here again the Sedona Principles build on the extreme position in order to avoid preservation. For example, even in paper discovery not every piece of paper was preserved. Indeed, only those pieces from people or portions of the organization believed to have relevant data to the

case were preserved. So Sedona accurately claims that not all deleted and residual data should be preserved while characterizing the preservation of relevant deleted and residual data as “exceptional cases”. When they are no more exceptional than any other data requiring preservation.

While the new rule changes recognize that all ESI may not be discoverable comments to the changes for Rule 26(b)(2) also recognize that,

“A party’s identification of sources of electronically stored information as not reasonably accessible does not relieve the party of its common-law or statutory duties to preserve evidence.”

The position taken in the comments to the new rule changes clearly take a more practical approach than the Sedona Principles which seem to favor destruction of ESI over preservation. In light of the rule’s reluctance to carve out portions of ESI for preservation, an understanding of ESI preservation procedures and methods becomes more important.

The preservation of ESI can be split into two separate components. The first is to identify the various sources of ESI and these can be many. The second is to develop a method for preserving the various types of ESI that have been identified. As discussed below the types of storage media will determine the best method for performing the preservation. But first, the media to be preserved must be identified.

Identify Data Sources

The first step in preserving electronic data is to identify the various data sources. The data sources are typically many as there are methods for handling them and should actually begin with the identification of people, places and issues that likely will be relevant to the case. Once those people, places and issues are identified it becomes a simpler matter to identify the associated sources for ESI. Just some of the items that should be considered are those immediately accessible to the individuals involved such as personal computers and related accessories. Next, other data storage locations that should be considered are those located on shared devices and network resources. Finally, the data held in specialty devices and applications like e-mail servers, content and records management systems, and applications databases should be preserved. Each of these data sources and the methods for preserving their data are discussed in the following sections.

Personal Computers and Accessories

Personal computers includes the individual computer workstation and/or laptop computers. The data to be preserved on these devices includes both their internal storage devices, which are usually hard drives but can include other technologies such as flash cards, as well as their

accessories.

The accessories to a computer include a variety of different media such as compact disks (CD) and digital versatile disks (also known as digital video disks or DVD for short). Although used less frequently these days, another type of external storage device is a diskette and ZIP disks which are a kind of super diskette. Newer to the technology spectrum are memory sticks, also known as thumb drives, flash drives and/or jump drives. These tiny devices, that are about the size of a disposable cigarette lighter, plug directly into a computer and commonly have as much capacity as a CD but can be found with capacity equaling a DVD. Also very common are external hard drives that connect to a computer in the same manner as a memory stick, by way of Universal Serial Bus (USB) or firewire.

As technology evolves the lines are blurring between personal computers and other devices. Some examples of this group include personal digital assistants (PDAs) and cell phones. These days such devices provide both their historical features as well as many others all bundled into one device. Still another example is the iPod that can connect to a computer and be used to store data files.

Shared Devices

After considering all of the personal devices one must start to broaden the net to consider many data sources that are not so personal such as the network. After all, some would say that the

network is the computer.

Virtually no computer stands on its own any more. Rather they are all interconnected to not only facilitate cooperative efforts but to share resources like printers, storage areas and other specialized services. Data sources for network data should include both on-line and off-line storage media.

On-line storage media are the readily accessible storage like the user directories and departmental shares on a file server or network attached appliance. The off-line storage media typically include archival and disaster recovery systems and their media. Although tapes are the most common form of storage for this type of data, other on-line backup systems, through providers like AOL, are also becoming very popular.

Specialty Devices and Application Databases

The next category of data source is the more specialized computer systems such as e-mail, content managers and document managers and application databases. Everyone is familiar with the e-mail server. It acts like the local post office for all the users connecting to that server. Users use their server mailbox to send and receive mail and often to store mail. Thus, with the prevalence of e-mail this data source becomes an important source of electronic data and one that is always high on the preservation list.

Enterprise content managers and record managers are designed to satisfy the organization's record retention requirements. They are typically useful in large organizations where uniform applicability of an organization's record retention policies across hundred or thousands of employees can be difficult. Thus, these systems enable a central administrator to push down to individual employee machines the rules that should be followed for retaining an organization's electronic data. The rules can include what to retain and what to discard, as well as where the retained documents are to be stored and organized.

Finally, there are application databases. These systems are found in almost every business application from financial accounting to human resources to project management and intellectual property. They essentially retain records related to individual transactions and provide management with an automated means for managing and reporting on this data. Historically, people have focused on the paper based reports that are generated by these systems. In recent times, however, they have learned that there is often far more data available in an application database than ever is presented on paper and that often times this data includes audit trail and identification features that are essential to an investigation.

Preservation Methods

Preservation is often quite simple and not as expensive or time consuming as many might

believe. The important part is to get the data preserved and then the more expensive and time consuming part, the analysis, can be performed in accordance with an appropriate plan.

This is a good place to remember that the preservation phase is not the place to cut corners.

Preservation is typically not that expansive or expensive. So, there is really not a lot to be gained by cutting corners during the preservation phase. Also, everything that comes later depends on how well the preservation phase was performed. Once it has passed, the data can never get any better.

During the preservation phase every case should be treated as if it will end up in court. It is easier to regard the computer as evidence from the start and ease up on the subsequent evidentiary analysis phase if it is determined that there is no substance to the issue. The opposite approach, however, is not impossible. So, it is best not to start working with the computer data in a casual manner and then realize that there is a problem. By that time it is often too late to start treating the data as if it were evidence. Instead, treat it as evidence from the start and practice good preservation methods. The particular techniques that should be employed will depend on whether the data to be preserved is located on a read-only device, a read-write device, within specialty applications or on archival and disaster recovery systems.

Read Only Devices

The easiest place to start a preservation effort is with the read-only devices such as CDs, and DVDs. These items are the easiest because they can be preserved simply by taking physical custody of them and storing them in a secure location. If they should happen to contain data that is needed, then it is fairly easy to make copies of their contents and provide a working copy back to those that need the data.

When making copies of these devices it is also good to remember that CDs and DVDs can contain multiple sessions. These sessions will not be visible when the contents of the CD or DVD are viewed through the Windows Explorer. It takes other types of CD viewers.

Evil doers wanting to hide their activities can use this fact to create new sessions on top of the data that they really want to hide. Consequently, to be sure and copy all the data on a CD one must perform a track-to-track copy. If, instead, the copy is accomplished using drag and drop features of Windows only the last session will be copied to the new CD or DVD.

Read-Write Devices

Things become more complicated when preservation moves to the read-write devices such as personal computers, PDAs, cell phones and memory sticks. After all, simply operating a computer or examining these devices without taking special precautions to protect their contents

can alter their data. Typically the dilemma is that these devices are likely needed for continuing business operations. So, it is unlikely that preservation can be accomplished simply by taking custody of them and storing them in a secure location. Thus, these devices will likely need to be copied.

The exact copy method that should be used can be guided by several factors. If the device is a personal computer there are several options. If the device is immediately needed for service then the simplest method is to perform a drive swap.

A drive swap can be performed by using special software, like Norton Ghost, to copy only the active data, including system files, from the original drive to the replacement drive. By only copying the active data, the copy process will be faster than if the entire drive, including free space, is imaged. Furthermore, there is no need to copy the free space containing deleted data. By deleting the data the user has already determined that the information is not needed for continuing operations. After completing the copy process, the original drive is taken into custody while the replacement drive is installed in the target machine and business operations continued.

Drive swapping is always an option that should be considered. The only issue is should the swap be performed after a forensic grade image is created and should the replacement drive be created from the forensic grade image. Some professionals may want to take a more cautious approach and not want to risk anything happening until after the forensic grade image is created. For

example, if the Ghosting process is accidentally performed in reverse then the original drive will have copied to it all the active data, if any, from the replacement drive. Also, some may not want to risk having the computer accidentally boot into an unprotected state where the metadata of files on the original drive could be altered.

Certainly, all of these risks are possible but there are methods for protecting the original drive from the effects of Murphy's law. First, disconnect the original hard drive from its connection while configuring its host system to boot from a floppy drive or CD drive. Next, perform the Ghosting with the original drive in the drive 0 (zero) position. Ghost will not allow images to be copied back onto the zero position. Finally, use the Ghost DOS based executable on a properly prepared DOS boot diskette or CD. In order for the DOS diskette to be properly configured remove any references to the original hard drive from the system and configuration files on the boot disk.

Drive swapping is not always practical. A court order may not contemplate drive swapping. Similarly, in enterprise situations it may be impractical from a timeliness perspective or from a notification perspective to perform a drive swap. Since a forensic grade image is considered the evidentiary equivalent of the original, there technically is no real need to retain the original. Rather, retention of the original is desirable only if questions later arise about the validity of the imaged copy.

In those cases when drive swapping is not practical, a forensic grade image is the next best option. A forensic grade image is a bit stream, sector by sector copy of the original. The image

can either be copied to a file or replicated/restored to another hard drive device. The latter process is often referred to as a clone, since it is a working copy of the drive. This is different than the former in that when the image is copied to a file, the image cannot be used to boot the computer. Nor is it otherwise functional. This state is frequently preferred over the clone version because of the file's increased resistance to alteration and increased protection of the imaged data that is encapsulated in the file format.

As with the working copy process described previously, special procedures will be required when making the forensic grade image to ensure that no changes are made to the original prior to completion of the imaging process. The image must be made in an environment that will not alter evidence on the original drive or it must be performed with the assistance of write blockers that protect the original drive from alteration.

Creating forensic grade images can take more time than drive swapping. The increased amount of time is caused by the technical difference in the process as well as by the additional assurance steps that are now practiced. As explained previously, the drive swap will copy only the active data while the forensic grade image will identically copy all data on the hard drive, including the free space on the drive. Additional time will also be required for verification of the forensic image. Even when tested and reliable imaging tools have been selected for making the image, current best practices require that the image be verified by at least one of two options. The first is by re-imaging with different equipment and then comparing the MD5 hash, or equivalent, of the first image to the MD5 hash of the second. In the alternative, one could also compare the MD5 hash or equivalent of the original drive to the MD5 hash of the image. Either method will

require making at least two passes over the original drive.

An MD5 hash is a one way algorithm that computes a unique value for a data stream like the contents of a file or the contents of an entire hard drive. The MD5 hash is but one of many such algorithms. The MD5 hash was developed in 1994 for use in electronic signature authentication. It has been widely used in computer forensic applications as a means to determine whether two data streams are unique or identical

In recent times the MD5 hash has come under attack, since it has been determined that collisions are possible, particularly when the two data streams are very short such as a date. A collision is where two different data streams produce identical MD5 results. As the data streams increase in length and complexity, however, it is believed that the chance of collisions diminishes.

Interestingly, the MD5 hash was never incapable of experiencing collisions. Rather it was just believed that with the chances at less than 1 in 2 raised to the 128th power that collisions were very unlikely. (2 raised to the 128th power is about the same as 3.4 times 10 to the 38th power or the number 34 followed by 37 zeros. Thus, the results produced by a MD5 hash are statistically far better than those experienced in DNA analysis.)

In any event, performing the verification can take considerable time and this would be in addition to the time required to perform the original image. Naturally, if the verification process does not produce identical results then another imaging or verification attempt will be necessary.

Thus, it is easy to see how the imaging process can take far longer than simply performing a drive swap and retaining the original. Of course, at some time the imaging will have to be performed but if the goal initial goal is to reduce the cost of preservation and to reduce disruption to the organization then performing a drive swap in the field is a more efficient alternative than performing a drive imaging in the field.

Other than computer hard drives there are several other types of read-write devices that will need special processing in order to properly preserve their data. Memory sticks, cell phones, digital camera, PDAs, digital voice recorders, and all other kinds of digital recorders will also need special processing. The precise manner in which the data on those devices can be preserved will depend on the type of device and the manner in which one can connect to the device.

Memory sticks can usually be imaged similar to a hard drive but a special write blocking device matching the memory stick's connection is needed. Digital cameras and the removable memory chip in cell phones are similar to memory sticks but also require a special device for reading their flash memory chip in a protected mode.

PDAs, voice recorders and the embedded storage devices of cell phones can be more troublesome because they typically require special imaging tools. In other words, the same tools for imaging hard drives and the other devices previously discussed cannot be used on PDAs, voice recorders or the embedded storage of cell phones.

After all of the read-write personal devices have been preserved, the next storage category is the shared storage devices like network servers and other storage devices. Network storage devices introduce another set of issues. One of the evolving problems is the sheer size of network storage devices. In recent times the storage capacity of these devices has become very large. As a result, imaging can take many hours and even several days to complete. Thus, performing the image while minimizing business interruption can require careful scheduling.

Another complexity introduced by network storage is the configuration of storage media. Often the media is assembled in a RAIDed configuration. RAID stands for Redundant Array of Inexpensive Disks. In other words the large capacity of the network device is not accomplished with a single large capacity hard drive but by taking many hard drives and configuring them so that they operate as if they were a single device. This configuration can mean that individual drive imaging is not practical. As a result, a different approach is required and usually results in imaging the entire array as a logical device instead of a lot of physical devices.

Other than the technical challenges, the particulars of network storage also introduce some practical challenges as well. In other words, if the total data storage volume is large but the area of interest is small, such as a single directory or folder, the question arises whether imaging the entire storage device is even warranted if the information of interest can be obtained from other data sources such as backup tapes. Since the server is not being used by an individual some of the things that one would expect to find in the free space of the drive on a personal computer (such as application temporary files that are created when one views a document or link files

pointing to the location of the document or internet history files) will not be found on the shared network device. So, if a good backup tape history exists one might not expect to find much more on the network device than what would exist already on backup tapes. If not, one might choose to simply take the key artifacts such as files as they currently exist, access logs and other user related information.

To what extent a server should be preserved can also be based on some analytical testing. Signature and hash analysis are means for evaluating server contents. Hash analysis can be used to examine active files to determine whether or not the data being sought exists. Similarly, individual sectors or clusters can be hashed of either allocated or unallocated space and compared against similar segmental results from data of interest to determine whether it ever existed on the server.

Also, signature analysis can be performed against free space to determine whether the content of free space contains data for the types of documents that might be of interest.

Specialty Applications

E-mail and other specialty applications such as databases are the next area of preservation interest. For these items, it might not make sense to preserve the entire device on which these applications reside, if all that is of interest is a smaller set of data within the application itself. E-

mail is a good example of this condition. What is the practicality of preserving an entire storage device if all that is of interest is a few mailboxes within the e-mail data store? This is particularly true when the e-mail system encapsulates all of its data within a database file. In such a case there will not be a lot of remnants of deleted e-mail scattered around the storage device, particularly when the device is only a storage location and is not the viewing location of the e-mail data. Thus, in these cases it may be just as suitable to preserve the e-mail data store along with the individual mailboxes of the persons of interest. There would be little else to gain, if anything. Preserving the entire data store along with individual mail boxes would be desirable because any deleted e-mail that is recoverable will likely not be tagged to the individual mailbox even though it remains within the data store. So, preserving only the individual mailbox would not get deleted but potentially recoverable e-mail.

When preserving the data store, it is usually unnecessary to preserve anything other than the data store. There are tools that can read directly against the data store without the other data contained on the server. So, when it is time for the examination to be performed, there usually is no need to rebuild the entire server. Of course, when planning the discovery and preservation one should know what data the examination tools actually require.

Content managers and records managers provide another specialty application. These programs help an organization to retain and manage its record retention policies. They may or may not actually store the data. Rather, they may only point to the location and organization of the data that is stored elsewhere. If they do, in fact, store the data then one would want to determine what

data is available and the best way that it should be extracted from these system and preserved.

Application databases also provide a special set of issues. Most notably is that there may not be any real need to image the storage devices on which they reside, which could be quite large, if an adequate history can be obtained by preserving historical backup tapes. At most the entire database file will need to be preserved but at a minimum there may be ways to select only the desired data from all of the data that is contained within such a system.

Documentation

In addition to performing the preservation it is also important to begin documenting the chain of custody of the preserved data. This should begin by inventorying the items and recording the date on which they are taken into custody along with any unique identifying marks such as serial numbers, etc. When the items are taken into custody it may not be a bad idea to take photographs of the data, particularly if they involve the contents of a person's office. In that case it is good practice to take pictures of the office and record as part of the preservation effort exactly where the item was found. This extra step will be helpful later on when the dispute is exactly what was found, where it was found and why something else was not found that someone claims should have been there.

Conclusion

In summary, the new rules have not altered one's duty to preserve ESI. Unlike the Sedona Principles, the new rules do not favor destruction over preservation for any particular class of ESI even when it is inaccessible data and subject to the various discovery exceptions.

Fortunately, preserving data does not have to be expensive, time consuming or burdensome. It only needs to be preserved. The examinations can be conducted later. So, in many cases it simply is a matter of pulling the ESI aside and taking it out of daily use. In the remaining cases there remains several options for minimizing the cost and impact of the preservation effort.

The first step is to identify the people, places and issues relevant to the case. Then identify the ESI that is associated with each of those people, places or issues by identifying the various storage locations, media and machines that they are able to access.

Once the ESI locations and media are identified the next step is to preserve the data. There are a lot of methods that can be used. Choose the one that properly preserved the complete spectrum of data and provides for the most efficient and least disruptive method.

1. *Mosaid Techs, Inc. v Samsung Elecs. Co.*, Case No 02-C-6832, 2003, WL 22439865, “While a litigant is under no duty to keep or retain every document in its possession, even in advance of litigation, it is under a duty to preserve what it knows, or reasonably should know, will likely be requested in reasonably foreseeable litigation.”

2. Rule 401, Federal Rules of Evidence.

3. *Op cit*, “negligent destruction of relevant evidence can be sufficient to give rise to the spoliation inference, if party has notice that evidence is relevant to an action and either destroys or allows destruction by failing to take reasonable precautions; ‘willfully blinding’ oneself, failing to place a litigation hold on e-mails permitting automatic deletion of e-mails on a rolling basis, and complete and utter failure to produce e-mails responsive to document requests prejudiced adversary, threatened the integrity of the court and warranted giving of adverse inference and sanctions of more than \$500,000.