

Stop the Bleeding: Cost Containment for Electronic Discovery

By Sharon D. Nelson Esq. and John W. Simek

© 2009 Sensei Enterprises, Inc.

There is one sound always associated with announcing the estimated costs of electronic discovery. It is a gasp. Sometimes accompanied by a wheeze of pain, sometimes by a grunt of reluctant acceptance, and sometimes by a curse of protest. No matter how you slice and dice it, electronic discovery is not cheap.

Still . . . inquiring minds have wondered whether it really needs to be quite as expensive as it is. Technology is a marvel – can't we employ technology as a friend and ally in cutting discovery costs? Cost-cutting in EDD is about both people and technology. We will speak of both, but focus primarily on technology.

Let us set the stage:

You just received a letter from a competitor threatening a lawsuit, claiming that a recent hire has stolen proprietary information and brought it to your business. Perhaps you didn't actually receive a letter, but you have reason to believe that there will be litigation involving your company. Either way, you are now in a litigation hold and must quickly preserve any evidence pertaining to the alleged wrong doing.

Larger corporations probably have records management systems in place to deal with the rapid identification and preservation of information. But not everyone is a Fortune 10 company with huge budgets to implement data preservation systems just for the purpose of preserving evidence for a case.

The more common scenario is for companies to have a rudimentary system of information categorization and retention that doesn't lend itself very well to reacting to the litigation at hand. The initial reaction is to "save" everything just in case it may be relevant. After all, storage space is cheap and getting cheaper. This initial premise will significantly raise your costs for the litigation. The single, most costly component of electronic discovery is the attorney review. Therefore, the goal must be to keep the data volume to a minimum.

Where's the data?

Hopefully, you have some sort of records management system that can aid in identifying what data should be preserved. If not, then you need to develop a data map that identifies the potential location of electronic information. Don't forget data that may be held with third parties. As an example, if you outsource your payroll and the lawsuit involves employee compensation, then you need to notify the payroll company not to destroy any information and to preserve data pertaining to compensation.

What if you don't know where the information is and you haven't restricted where employees can place data? As a minimum, your computer usage policy should identify the acceptable uses of computers and what equipment is acceptable. USB flash drives are a dangerous source of data theft if not properly

controlled. All too often disgruntled employees will copy corporate data to a USB flash drive as they depart their employment. You could lock down the USB ports on the computer, but that may hinder real business usage such as digital cameras or USB printers. Your first technology tip is to install USB monitoring software that can centrally log all USB activity on every computer.

What if you still don't know where there data is and you never mapped out all the potential locations? There still may be hope for you. With the increased interest in electronic data discovery, there are now technical solutions to aid in finding the data. You should purchase a hardware device that connects to your network. These devices work in a fashion similar to a web crawler such as Google. The device "crawls" the network and identifies files, owners, contents, etc. Make sure that the device will work in your environment. As an example, it may require that file and print sharing be enabled on all computers or that the crawling agent be given administrator privileges. Some manufactures price their devices based upon the amount of data "discovered" and others price the device based upon the amount of volume it can store.

One step many organizations take to be more proactive is to implement content archiving software. These tools archive and manage the high-volume, user-generated content like email, files on network shared drives, and SharePoint sites that tend to be not otherwise centrally managed. IT departments like archiving because it reduces storage costs and makes it easier to backup their systems. Archiving also benefits organizations by providing a central repository for information to deduplicate, manage retention, and conduct eDiscovery searches and legal holds. "Given that e-mails and network file share documents are the two most requested types of content in eDiscovery, we've seen a huge demand for our unified content archive that supports email and file system archiving," says Barry Murphy, Director of Product Marketing at Mimosa Systems. "The fact that they can set very granular retention policies and item-level legal holds on this extremely difficult to manage content really helps them to sleep better at night."

Active data only?

Another cost consideration is whether only active data is at issue. Active data is the data that the user can see when operating the computer or device. It does not include any deleted (and normally irretrievable) files. Forensic processes are needed if you must include deleted data or data that may reside in unallocated space on a hard drive. Forensic acquisitions are expensive and typically not needed. Many times we see the attorneys jumping up and down wanting to forensically image computers to preserve ALL of the data. In a typical company environment, this can be overly expensive even if done by internal IT personnel, which is generally not a good idea for the reasons stated below.

If forensic images are required, only do the key players, if you can secure that agreement from the adverse party. We would recommend that the forensic imaging be done by a third party and not your IT staff. This is not only because of the vested interest, but creating forensic images is not just copying data or plugging a disk into a piece of hardware. There are set procedures and verifications that are performed as part of the imaging process, which most IT personnel are not familiar with.

So how do you minimize the business disruption and keep costs down if you really do need forensic images? A good practice tip is to have your IT folks clone the appropriate drives, allowing users to run off the cloned drives while you secure the originals. This will save you a lot of money since you are not initially creating the forensic images. IT personnel are familiar with using the Ghost (Symantec) product to create drive clones. Ghost is not a forensic tool, but can be used to create a logical copy of the user's hard disk so that they can get back to business. Securing the original drives means that they are available for any future forensic analysis if needed – and often, it is not.

When would you need forensics? Forensic analysis would be used if web-based mail (e.g. Gmail, Hotmail, etc.) is at issue or if locally stored files from the user's computer are at issue. Typically, employees using web-based mail aren't aware that the Internet "trail" is stored on their local hard disk. They may even purge their Internet history in an attempt to remove any trace of web-based mail activity. Computer forensics can still recover this data and present the Internet activity of the user. Computer forensics is also used where data theft is suspected. Normally, we see insertion of USB devices and copying of data to those devices. These are just two examples of where computer forensics may be utilized. The good news for your pocketbook is that most cases do not need forensics and dealing with the logical data is sufficient to locate and preserve the relevant information.

Metadata Preservation

Sometimes you will need to preserve the system metadata related to the electronic files. This is different than the metadata that is imbedded within the file itself. The system metadata would be such things as the file creation date or the file accessed date. Merely copying files to another location will modify both of these values, thereby not providing a true representation of the original data. There are tools that can be used to preserve these values if needed. Arguably the two most commonly used tools are ROBOCOPY (command line Windows tool) and SafeCopy 2 by Pinpoint Labs. As with all tools and handling of electronic evidence, make sure that you test the software prior to actually dealing with real evidence for the case.

De-duplication

Why would you need multiple copies of the same information? You wouldn't, but how do you know if it is the same or not? Reducing data volume means reducing reviewing and production costs so de-duplicating is key to your data processing. It goes without saying that your records management system should be dealing with this up front, but what if you didn't address it or you just don't have a system to manage your records yet?

The obvious first question is: what constitutes a duplicate? You could use a hashing algorithm like MD5 to identify exact duplicates. This will work admirably for your electronic files like spreadsheets, documents, etc.

Be careful when processing e-mail though. Does your tool hash an entire e-mail message including the attachment? De-duplication of e-mail messages is not as easy as it sounds and is better outsourced to those that have specific experience and tools. More often it is appropriate to remove near-duplicates,

which have the same attachments, recipients, subject line, etc. but have different MD5 hashes due to the message ID and seconds time difference in the headers.

Finally, you want to remove any known system and program files. Typically, you use a database of hash values to identify these innocuous files. The industry tends to term this deNISTing because the NIST (National Institute of Science and Technology) database is used to identify the known files. The NIST listing contains over 34 million hash values for operating system and application files. Some commentators claim that the filtered files can be in the range of 30-50%, thereby significantly reducing the review files.

The End Game

The de-duplication, subsequent electronic data discovery processing and production will probably be accomplished by an EDD processing vendor, although many of the very large corporations are beginning to bring this processing capability in-house. The key to successfully dealing with electronic discovery is taking reasonable steps and documenting your actions. Technology can help with the identifying and gathering of the data, but people component will be the ones to make decisions as to what types of files to preserve, which custodians should be addressed, what keywords or phrases should be used for searching, etc.

This entire process is not for the faint of heart. Few attorneys are competent in this arena. By in large, you are likely to be guided by your EDD expert, so choose that expert early in the game before you accidentally go astray. The monies you spend on an expert should be recouped many times over by the efficiencies of a sound process and the reduction of data volume, which will save huge sums of money in reviewing and processing costs. Truly, that's advice you can take to the bank.

*The authors are the President and Vice President of Sensei Enterprises, Inc., a computer forensics and legal technology firm based in Fairfax, VA. 703-359-0700 (phone)703-359-8434 (fax)
sensei@senseient.com (e-mail), <http://www.senseient.com>*