

Electronic Document Discovery / Litigation Forensics

**Prepared for The Institute of Law Clerks of Ontario
September 29, 2004**

**Girts Jansons
JLS inc.**

**JLS inc., 347 Bay Street, Suite 1100, Toronto, Ontario, M5H 2R7
1-800-979-9139 www.jls.ca**

Definitions

Electronic Document Discovery is the term used to describe the collection methodology and production procedures used in gathering documents from files generated by and/or stored within computer devices for use in civil discovery. This includes *active files* which are represented by the current undeleted electronic files on your computer system.

In this paper, we will refer to *Litigation Forensics* as the extraction and use of hidden information, called metadata, embedded in active files electronically. *Metadata* is information intrinsically created by computer programs or operating systems to identify certain useful data specific to the file such as the identity of the creator, the date the electronic document was created, the date the document was modified, to name just a few.

Litigation Forensics should not be confused with *Computer Forensics* which is the thorough examination of computer hard drives for evidence of fraud, spoliation, or criminal wrongdoing, such as child pornography. This process examines not only the active files and metadata, but also latent data and data located in slack space on the hard drive. *Latent data* consists of deleted files, memory dumps, swap files, printer spool files and other data that is not visible to the end user. The recovery and examination of latent data usually requires specialized computer utilities and the expertise of highly trained certified forensic specialists. These individuals are typically former police officers who have devoted years to working within a forensics lab. Computer Forensics can be a very time consuming and expensive proposition, even with the most sophisticated computer tools and forensic experts.

The Difference Between Paper and Electronic Evidence

Paper discovery has been the norm for many years and, thus, the processes associated with same have been perfected. Litigators are familiar with the collection of paper documents and most clients will have sound knowledge as to the relevant sources of information and where the corresponding important paper documents are being stored.

Essentially, the gathering of paper is a straightforward sweep of all the documents that may be considered relevant to the issues of the lawsuit. At this point, two options exist in contemplation for discovery: (1) leave the documents in paper format and categorize the materials in a word processing document or database; or (2) scan the paper converting each page to an electronic image that is linked to a matching record in a database.

Regardless of the method selected, each document must thereafter be reviewed to determine actual relevance or privilege. Then, if left in paper format, the

relevant documents usually get assigned a document number for reference purposes and the collection is photocopied at least twice for production. If the documents are scanned and linked to a database, the allocation of a Bates numbering convention for the responsive set is seamless and the creation of a list quick. The documents themselves can be produced in hardcopy by printing directly from the database, or electronically by burning the TIFF images to CD.

Electronic document discovery, on the other hand, is considerably more complex. Although it has been a serious reality for some time now, many litigators still do not seek and/or request all this discoverable evidence. Today, just asking for paper evidence will render a very small portion of the available material to discover. It is estimated that 90% of most companies' information is stored in electronic format, with the moderate volume being approximately 55%. While paper may constitute a portion of this collection, pursuing the electronic data will reveal evidence missing from the paper set that may decide your case or devastate that of opposing parties. Now, more than ever before, it's likely that the smoking gun will be found in electronic format.

Electronic data provides much more information than paper, such as the metadata. In addition, it's simple to make instant changes to, provide comments about, create formulae within, embed pictures, links or other files, associate notes into and search the content of a document while the document is in electronic format. It's also extremely easy to distribute the electronic documents simultaneously to a large number of recipients including parties who may not necessarily need to review the material but are receiving it for informational purposes only. It has become far easier and faster to communicate by sending an e-mail, rather than picking up the phone, whether for business or personal purposes. This communication via e-mail extenuates the creation of electronic files and, since it assumes characteristics similar to a discussion, "dialogue" may be recorded in a written format that would not otherwise exist if limited to the telephone conversation.

There are many reasons to, and circumstances that, encourage the creation and retention of electronic documents. Firstly, paper requires extra steps to generate and is difficult and cumbersome to maintain. Secondly, with the cost of paper storage space increasing annually, the trend is for companies to destroy their paper or convert it to electronic images. Further, with the capacities of hard drive space continually increasing and the costs for same decreasing, companies and individuals are more likely to save excessive amounts of electronic data than truly necessary, even if the retention of any given document is purely intended as a "CYA" safeguard or as an emergency backup.

Electronic – how much is there?

There is an estimated 90 million electronic pages created every day world wide, and a small portion of these get printed and converted to paper format. Virtually every establishment today not only utilizes computers to perform most tasks to efficiently and effectively conduct various business activities, but also incorporates computers as a means of communication both internally and externally. Sending documents by snail-mail is considered far too slow to accommodate the needs of today's businesses. The preferred method of communication is e-mail, considering the ease within which e-mails can be created and exchanged, and also due to the fact that an abundance of matters now require immediate attention necessitating instant collaboration and exchange via e-mail. The initial "electronic document" often results in a series of e-mail exchanges not only with the original parties but an ever-expanding list of people.

If in a company of 50 employees, each employee deals with 30 business e-mails a day, the volume per year totals 375,000. The likelihood is that many individuals send and receive far more than 30 e-mails per day, therefore, 375,000 is a modest reflection of actual e-mail activity. This does not include bulk distribution, spam, advertising, personal exchanges or attachments. All these, of course, would substantially increase the total volume of electronic files relating to E-mail alone.

With hard drives reaching the 300 gig size today, massive amounts of data can now be stored on a single hard drive. To do some calculations, let's use some media that we should all be familiar with:

- 1 kilobyte (KB) = 1000 bytes
- 1 megabyte (MB) = 1024 kilobytes
- 1 gigabyte (GB) = 1000 megabytes
- 1 terabyte (TB) = 1000 gigabytes

Further:

- the 3 1/5 inch floppy = 1.2 MB and will store about 100 pages of electronic documents
- the popular CD = 640 MB and will store about 50,000 pages of electronic documents
- the DVD = 4.7 GB and will store about 375,000 pages of electronic documents

- a 300 GB hard drive will store about 24,000,000 pages of electronic documents
- a data server with a terabyte of storage space will store about 80,000,000 pages

Note that volumes may exceed these numbers if some form of data compression is used.

As you can see, as storage space gets bigger, cheaper and more accessible, the quantity of electronic documents grows exponentially and, thus, it becomes hard to ignore the availability of stored electronic data as evidence.

Electronic – where to start?

It is assumed that each party and its respective client will produce all the relevant documents whether originating in paper or electronic format. So, in the quest for electronic documents and with the vast amount of electronic data available to discover, it often becomes a matter of where to begin.

Considering that electronic documents possess unique information housed in the metadata, it is essential that the collection process be done properly to ensure that no information is altered or lost. Therefore, adopting a proactive approach in terms of the preservation, recovery, retrieval, treatment, review and management of electronic discovery is a worthwhile and valuable commitment to make to avoid emergency procedures that will undoubtedly result in increased costs and perhaps inferior results.

Once there is any indication of impending litigation, the plans for preservation and collection should immediately commence. These terms are distinctly different. If ill-informed, many clients will want to embrace aggressive tactics, with the sincere aim of assisting counsel in their collection efforts. They'll eagerly start opening, closing, deleting or moving files. They'll forward relevant E-mails either directly to counsel or to a central server intended for transfer to counsel. All these activities substantially modify the metadata, rendering most of it useless for litigation purposes. Then, if the original files are not resurrected as a result of these actions through computer forensics, counsel is left with the burden of claiming privilege over many of the documents or redacting a large amount that would otherwise have been entirely responsive in nature. So, in the preservation of electronic documents, there is a need to educate clients advising that the electronic files be preserved in their original state and location.

Who should be identified in the collection process?

The amount of actual electronic data targeted for collection should be governed by the number of custodians thought to possess relevant material. A first step

would be to conduct an initial client interview to determine potential document sources and then investigate those. Determine which individuals clearly possess relevant data and circulate preservation notices to those custodians. A preservation notice to your own client should be informational in nature, providing guidance and instructions to assure that no data is modified and that the collection procedures are adhered to. Ensure that you receive an acknowledgement to your preservation notices. Upon learning that certain anticipated sources do not, in fact, hold relevant data, consider a protective order granting your client the ability to continue administering its standard retention policies.

In addition to the relevant custodians, who else may be requested for examination where the case involves electronic document discovery?

The relevant custodians will not likely possess all the information necessary to respond to questions dealing with electronic document discovery. It is wise to make a prospective list of individuals in the IT department or elsewhere who have working knowledge of the mapping and management of computer systems. Preservation notices and replies should be exchanged with all persons identified through this process including advice not to recycle backup tapes. The technology staff will further be able to assist with the following questions:

- What are the network configurations?
- Which servers have relevant information on them?
- What operating systems are on the servers?
- Are there any snap or backup servers? Some systems run mirror servers so that if one fails, the other is available to avoid business disruption.
- Are there servers in different physical locations?
- Are there any sub-networks or department specific networks?
- Are there any standalone systems?
- What software is used for e-mails and business?
- Do different departments use different software?
- What are the current software versions as well as previous ones?
- What are the current backup procedures and policies and are they being adhered to?
- What backup software is being used including versions, tape types and size and/or other backup media?
- Is there a log, index or cataloging of the backups? This helps determine how much may need to be restored.
- Are the backups incremental or full?
- Who performs the backup processes?
- Where are the backups stored?

- What are the current retention and destruction procedures and policies and are there any logs maintained to ensure that these measures are adhered to?
- Who in IT has worked on the systems in question?

Electronic – where to look?

An effort must be made at the beginning of collection to determine where information relevant to a case may reside. During the questioning of relevant custodians, learn what kinds of electronic files are created, what software programs are incorporated into their routine, and how the electronic files are disseminated and stored.

The electronic files will be located on the servers and computers used by the relevant custodians including mobile computers, home computers and individual desktops. These may be physically located in multiple corporate offices in different jurisdictions, depending on the nature and size of the litigation. Remember that most of the servers and business computers get backed up on a scheduled basis so when collecting, although the information may not be on a server or computer because it has been deleted, it may reside on one of the backup tapes. Keep in mind to check whether there may exist some historic or legacy data that may be within the scope of discovery. Electronic files also may be found on the servers and backups of the internet service providers (ISP) used by any e-mail author or recipient.

While business documents and e-mails represent the more common form of electronic collection, there are yet other forms of electronic files that may require investigation. The phone server that takes an incoming message creates electronic files that may become the subject of discovery. Cell phones have text messaging and phone books saved in electronic format. Blackberries and other PDAs all store information electronically. Then there's the digital cameras and similar devices that incorporate the use of many different types of memory cards, all of which can store electronic data.

Electronic - what will be collected?

In simplistic terms, everything that may be considered in any way relevant to the litigation, including all existing drafts and versions, should be preserved and collected.

Should there be a need to maintain data past the historical date range identified in the case, ensure that there is a systematic approach undertaken to determine, for example, who is continuing to create potentially relevant data and to develop controls that can be adopted to distinguish and preserve truly responsive materials for subsequent collection and processing.

If there is evidence of fraudulent or criminal behavior, then the search for relevant material takes on a different flavor. In these types of situations remember, as a rule, that no one wants to lose or have to recreate data that has taken time and effort to generate or collect. Although it may seem that the evidence has been destroyed, usually there is a copy to be found someplace that can be restored. This can include locating files hidden inside other files as well as many other computer tricks. So, the investigation becomes a little more complex requiring a creative experienced and certified forensics examiner to assist and, on occasion, a court order to demand search and seizure measures.

What data can be gathered from electronic files?

Most electronic files carry with them some form of metadata. In Windows 95 and 98, only Microsoft Office files had metadata attached to them. In Windows NT, 2000, and XP, the operating system attached metadata to each file. This information in many cases is invisible to the user but can be viewed through the properties of the file and otherwise extracted using special computer utilities. In e-mail, the most commonly collected metadata is:

- Authors, recipients and individuals receiving either a copy or blind copy of an e-mail
- E-mail addresses
- Date and time e-mail was sent or received
- Subject, or re: line
- Contents of e-mail body
- File name of attachment(s)
- Unique system ID number for e-mail
- Internet header information for e-mails sent externally
- Mailbox folder name in .pst file where e-mail maintained and unique system ID number for folder
- Unique system ID number for source of e-mail data
- E-mail status (read, forwarded, deleted, etc.)
- E-mail priority
- E-mail flagging
- E-mail size
- HTML content

In other electronic documents the metadata available includes:

- File name and path
- File type
- Software program
- File size

- Date created
- Date modified
- Date accessed
- Attributes
- Title
- Subject
- Category
- Keywords
- Templates
- Page count
- Word count
- Character count
- Line count
- Paragraph count
- Scale
- Links
- Comments
- Revision number
- Date last printed
- Date last saved
- Total editing time

E-mails

The considerations to be given to e-mails exceed those relating to electronic documents in other native formats, primarily because collection may be highly duplicative. A simple search will not give rise to the entire assortment and each contains valuable hidden data like bcc routing information not otherwise evident on certain versions. Other invisible components include activity logs, reply history, address aliases and contact information. More often than not, an entire e-mail account will contain a small quantity of relevant communications, leaving the bulk as non-responsive. All of these traits command the need for special handling of e-mails to ensure that only the relevant materials get produced and the appropriate metadata is preserved.

Electronic file issues

Some electronic files differ in appearance and content from their printed version, such that special attention and thought may need to be given as to the method of production. For example, electronic spreadsheets may contain hidden data in compressed rows and/or columns, formulae, white fonts and hyperlinks that will not be present on the printed edition. Some electronic files may be compressed for archival purposes requiring decompression to deal with the embedded documents. Others may have been password protected or encrypted.

Electronic – how will it be collected?

As with the preservation of electronic data, a detailed plan of collection should be implemented and followed using the appropriate tools, technology and, most importantly, personnel. Assemble a team that includes counsel, client representatives, IT staff and litigation support experts. The team will formulate a schedule of events to effectively identify documents, collect files, preserve metadata and track the chain of custody, and then assign tasks to the suitable individual(s). Determine from IT personnel if the resources and expertise exist internally. Establish how you will want to preserve the original data. You may need outside expert help if there is a need to deal with historic or legacy data, or with recovering deleted digital information. It may also be necessary to involve database administrators if your client uses workflow databases, financial databases, any data repository or specialized engineering software as these may not produce useful information without the assistance of the individuals who manage them.

An important objective during electronic document collection is to minimize the amount of intrusion on clients. The collection team should perform their activities as invisibly as possible, taking advantage of the server room to access centrally located files and to retrieve documents stored on individual computers through the network system. To further avoid disruption, it is sometimes prudent to collect whole accounts, drives and folders rather than concentrating on individual files. The individual documents can then be treated in isolation once gathering is complete with the consequential culling of irrelevant or duplicate material.

Electronic – how to cull to make volumes manageable?

Of course, if the electronic document collection is manageable, one may opt to review all the data, although, for extensive collections, a full review is neither warranted nor feasible, resulting in the need to minimize the collection to be reviewed.

The first step to culling is to identify the relevant custodians. The next step would be to identify the relevant time periods, typically determined by counsel and/or upon review of the pleadings. These two steps should isolate the bulk of the electronic data. If appropriate, the quantity of electronic data harvested may be further reduced by classifying relevant types of data, such as e-mails, spreadsheets and financial records.

The beautiful thing about most electronic data is that unlike paper, you can use technology to search its content to assist in culling. While one first imagines that relevant documents may not be caught, the use of key word searching has

become one of the most significantly accepted developments in electronic discovery as a way to manage the cost and size of documentary review.

The list of key words should be developed by the litigation team taking into account the issues of the case, names, e-mail addresses and phrases evident in paper documentation. Care should be taken to ensure that the list is not overly broad or, conversely, far too narrow. Incorporating the use of fuzzy and proximity searches, together with other sophisticated tools, will greatly reduce the volume of data to be reviewed. Multiple independent key word lists can also be generated to divide, for example, the responsive productions from the privileged set.

Another method of culling electronic document collections relates to the removal of files that are directly software related, such as system and execution files that are inadvertently caught through collection efforts. A further technique of volume reduction is the filtering of duplicates through the use of an algorithm to identify unique content. With e-mails, the quantity can be reduced through the elimination of spam, advertising, personal exchanges and jokes. If the volume of data is huge and spans years and years, if the parties agree, sampling of the electronic files and backup materials will also assist in the possible diminution of the volumes to be reviewed.

Electronic - how to produce – paper v. TIFF v. native?

Every litigator has their own thoughts on how they intend to produce their documents whether from a strategic, financial or other point of view, but it helps from a planning perspective to know at the start so that the suitable arrangements can be made. Questions to ponder include:

- Are you going to print the electronic files to paper and supply a numbered index?
- Do you want to supply any metadata as extracted?
- Are you going to convert the native files to TIFF images to lock and preserve the information as it existed at collection time and produce them with an associated database or text file?
- Are you going to supply searchable files with the content of the native files?
- Are you going to send the actual native files to opposing counsel? Remember that if you are going to deliver in native format, you will be supplying opposing parties with all the corresponding metadata that exists with each produced file. In some situations, this could lead to embarrassment, particularly if the file was authored by someone other than the individual claiming ownership.

Whatever method is preferred, one should always make an exact duplicate and preserve the electronic data as originally collected to respond to future requests.

How quickly can electronic data be destroyed?

Electronic data is perhaps more easily destroyed than it is created. The most innocent form of destruction occurs when individuals simply clean up their computer. As companies strive to become more efficient by upgrading their computers, servers and software, much data gets lost and compromised. Even as employees get hired, promoted, relocated or leave, data is lost as their laptops get reassigned, computers get upgraded and server space gets reallocated. These situations usually lead to the cleaning of drive space which may contain relevant data.

While many businesses keep backup tapes for years, others use them as the word implies - as backup – and, therefore, the tapes are recycled quite regularly. As the recycling occurs, the new backup overwrites older, possibly relevant data. Further, with the increased amount of electronic data, most companies have implemented retention and destruction policies to assist in curbing the amount of data required to be stored. These policies can lead to the destruction of valuable relevant data.

As employees work on existing and create new electronic files daily, they too inadvertently damage and/or lose valuable data through the unintentional overwriting of latent data on the hard drive which is still available to recover with specialized forensic tools. Latent data destruction becomes a serious concern in certain cases requiring the need to recover deleted files.

How deep do I need to go? In other words, should the computer drives be examined forensically?

In most situations, computer forensics is not necessary and, if undertaken, can release a completely different set of concerns. This level of examination differs from litigation forensics in that the former concentrates on data which has been deleted or lost by computer software, yet can still be found physically on disks and tapes. Computer forensics can locate and salvage information thought to be irretrievable. Simply because it's perceived that the data is not "where it should be" or it ceases to be recognized in a familiar format or seen by your operating system does not necessarily mean that it is gone. If there are no issues and supporting data, contemplated or real, of fraud, spoliation or criminal activity, then it is recommended that this process not be commenced for several reasons, the most significant of which are outlined below:

- The time to properly examine a hard drive is considerable.

collection efforts. It becomes immediately vital to send a demanding preservation notice to opposing counsel to assure that important evidence is not destroyed, either inadvertently or intentionally.

The clever litigators will use the guidelines and protocols they've implemented during electronic data collection to develop a line of questioning while examining opposing parties on discovery to determine whether suitable efforts have been made to produce electronic documents and, if so, to confirm that the methods were appropriately conducted. Since you have substantial knowledge, having gone through the many processes yourself, you should be specific in your targeted requests.

In closing...

In reality, much of electronic document discovery and litigation forensics comes down to common sense. Knowing that there is considerable information residing in electronic format that likely has not, and will not, make it to paper demonstrates a significant need to address this type of evidence. One cannot pretend that ignoring it will "make it go away". Rather, vigorous proactive planning is the way to attack electronic data, realizing that your case could very well be decided by embracing the tactical advantages that technology provides.