Electronic Discovery: How do you know you got everything?

By Virginia Jo Dunlap, Esq.

and

Robert Blatt, MIT, LIT

July 2009

Contact:
Blatt@eid-inc.com
(805) 529 - 0600
www.eid-inc.com

Contact:
Dunlap@eid-inc.com
(916) 838 - 6917
www.eid-inc.com

One of the first questions a litigator (or regulator) asks as soon as the electronic discovery response arrives is whether everything requested has been provided. Before electronic discovery, the hard copy materials would be reviewed, cataloged and indexed for easy future reference. If documents were missing, that fact might become evident through a content analysis or in a time-line review.

Not unlike the days when those boxes filled with documents arrived in the litigator's office, the electronic materials need to be treated similarly. In fact, because of the nature of electronic documents they may actually hold more clues than the boxed materials as to whether everything has been produced. However, finding those clues and evaluating them requires electronically stored information (ESI) forensics to be applied.

An early ESI forensics case

In 2002, we first worked together on some large regulatory investigations as part of a combined effort of multiple states and federal securities agencies. Jo was heading up the two California-led investigations of investment banks where analysts were suspected of having made public statements of support about certain stocks while privately disparaging those same stocks, often in emails. Although the first of these "analyst cases" was led by the New York Attorney General against Merrill Lynch and focused on emails, the bulk of that evidence was evaluated in hard copy form, using traditional investigatory techniques.

In the subsequent "analyst investigations", the bulk of the evidence was produced electronically. Not only did this generate an increase in the overall volume of information received but it enhanced the ability of the regulators to determine if what had been requested was, in fact, produced.

Robert provided the crucial technology support, conducting the ESI forensics necessary to establish that in one of our investigations, not all of the documents had been produced. Through a pattern analysis of the data it became clear that something was missing for several time periods. As it turned out, it was a lot of something missing. Additionally, it was clear information had been eliminated through a process referred to as "de-duping" to remove multiple copies of the same email. This "de-duping" process effectively destroyed the full communication threads that were necessary for a complete content analysis.

In our case, pattern analysis of all this electronically stored information consisted of three investigative techniques: 1) performing a detailed timeline analysis, 2) evaluating technical email metadata, and 3) verifying/building full and detailed message threads. The timeline analysis confirmed that we had not received all the documents. Through the technical email metadata evaluation we identified previously unknown mail and network servers where additional relevant ESI was stored. And, by reestablishing full message threads content analysis was more meaningful.

Combining our litigation and technology backgrounds gave us the advantage we needed to create a systematic approach to handling the voluminous document production. It also allowed us to develop a series of questions for key deponents about the storage and handling of the electronically stored information. Ultimately, the final settlement included payment on the part of one of the banks for the lax internal procedures utilized for producing the documents requested.

Since the "analyst cases", regulatory fines and discovery sanctions for failing to produce documents have become more frequent as companies struggle with the volume of data maintained in their systems. For example, the former investment bank Morgan Stanley agreed to pay federal regulators $15 million for failing to produce electronic records in several investigations over a period of years. This was after a civil judgment of $1.45 billion was awarded in a case against Morgan Stanley, which had not been permitted to introduce certain evidence as a penalty for the company's repeated failure to produce documents during the discovery process.

While the litigation support tools and technology solutions have advanced in the past decade to assist in tagging and sorting the content, forensic document and ESI analysis remains a cornerstone to understanding what has been produced and, too often, to what has not been produced.

Six ESI forensic steps to help answer the question of whether you got it all

Not every ESI forensic technique is applicable in every case or to every type of document or system. However, careful consideration of each of the following should be made:

- How was the material captured and stored originally;
- How was the information produced (ie: was it de-duped, was it produced as it appears in a single users email box, etc.);
- Creation of a time line;
- Pattern analysis;
- Analyzing the meta data and
- Analyzing the content.

Often, the nature of the ESI, the metadata contained in the documents and what evidence is being sought will guide the litigator (regulator) to the level of investigation needed in a particular case.

Understanding how the ESI was originally captured and stored

ESI forensics begins with understanding how the information was stored by the respondent, what format the data is in, and what sources of backup media need to be examined. ESI can be stored in email servers, instant messaging servers, websites, backup media, portable storage devices, network servers and desktop computers and home computers. ESI can also be stored in a system of software applications that have authoring and version control, routing or workflow controls or "paths and rules, along with a highly detailed audit trail.

Software applications referred to as Electronic Content (or Document) Management (ECM) technologies enable organizations to store very large volumes of ESI in any data format that can be stored on a computer. These data formats are almost always referred to as "native format" indicating they are stored in the format originally created by the user application. In contrast, ESI that is converted to a different format is considered to have been modified and altered from the "native format" through the conversion process. This almost always eliminates some level of metadata and at a minimum eliminates conversation threading, which is why it is recommended that ESI production require native formatted data in the structures created by the user

application (i.e., PST or NSF files). Understanding these technologies and how they work from the beginning of the capture process through how the information is stored is an important step toward answering the question of whether all the material requested has been produced.

An emerging issue to understand is whether the ESI was kept in a "trusted system" as described in statutes, regulations and best practice standards. AIIM has recently published guidance on what elements must be present in a trusted system and a similar statement is expected from ISO this year. If ESI is not maintained in a "trusted system" comprised of hardware, software, media with policies and procedures supporting it, then suspicions should be aroused as to whether the documents produced in response to a discovery order or subpoena have been altered in some manner from their original state. Our next article will focus on the subject of a "trusted system".

Information merging

Often multiple copies of electronic documents, especially emails, are produced and in order to reduce the volume of documents, a process of removing duplicate copies is applied by respondents. Documents that are retrieved from backup tapes are especially prone to having multiple copies, since backups may overlap in time frames. While removal allows for faster searches and reduces the number of documents to be read by legal staff, this raises serious issues.

"De-duping" or removing duplicate copies of content of an email may result in lost information. For example, if person A sends an email to person B and person C, and person B creates a new email (or forwards) this to person C, the removal process typically will result in only 1 copy of the email being received and the other copy is "de-duped". As a result the first copy of the email found by the "de-duping" software would be maintained and any other copies would be considered duplicate and removed. The problem that ensues is that if one of the parties receiving the email forwards to someone else, takes some type of action, or replies, the message thread linking this to a larger conversation is lost.

In the case of emails, often the same email appears several times – in the sender's box as well as in all of the recipients' boxes. While not technically de-duping, removal of the multiple copies may be appropriate, depending upon the facts of the case. If it is not important that someone received the email, then it may be sufficient to retain only the sender's copy and delete the others. In other instances, every version of documents may be kept in the final organized material. This process of merging relevant information results in the re-creation of full message threads and may also include extracting previously deleted messages from the mail file at the user level and the deleted message traffic performed by the server.

In the case of documents or records processed and/or stored in a document/ content management system not only do the actual documents or records become accessible, but also the history associated with any related transactions or actions becomes available for examination. For example, a review of the audit trail could assist in establishing whether documents that originally existed have been removed, renamed, merged with other files, who removed them and when. Other audit trails show who has accessed the document, performed updates, routed the information, and depending on configuration who has seen the information and provided input/feedback using workflow, email or other electronic communication methods.

It is through this approach to collecting and analyzing the ESI, either both mail systems, instant messaging systems and/or ECM systems that gives an inkling to a litigator [regulator] about what has been received have received and whether any obvious pieces of information are missing. Without this step, no one may ever realize that not all the ESI was produced.

Creation of a time line

In many cases, it is important to establish a time line of events and of documents related to those events. This time line helps establish what happened and when; it is built after all the information has been collected, merged as appropriate and analyzed. The time line should show when the message thread began and when it ended. In the case of documents stored in a content/document management system, the time lines will present when documents are received, stored, edited, reviewed, approved, finalized, distributed, etc. As such the value of receiving information in original format becomes very important for any type of full investigation where ESI is being requested.

Pattern analysis

Looking for patterns in the timeline associated with all the produced documents can provide significant assurance that all the documents were received...or not. In one of our cases, a pattern analysis of the produced documents showed that there were significantly fewer documents in certain months than in most others. That led us to initially question whether all the requested ESI had been produced. As it turned out, a lower level IT staffer charged with production tossed aside the tapes he was unable to read, weren't readily available or didn't think of requesting from the backup tape storage location. Had we not performed this type of pattern analysis we might never have discovered significant information had not been produced. .

Another type of pattern analysis involves comparing the volume of documents, such as emails produced by a particular individual during specific time frames. If that individual typically produces 1500 emails per month, yet one month falls off by 60% or 70% (or more) and there is not other reasonable explanation (i.e. the employee was on vacation or out ill), then questions may be raised about the reason for the decrease. This also holds true for the number of emails being received by an individual.

Pattern analysis can also involve following threads of emails to identify gaps in traffic, associations with other communication methods and most importantly how the user organized the information. The same is true when individuals store and index ESI on ECM systems which also may incorporate full text search and retrieval capabilities of not only the metadata but also content within the documents or records. Most people organize electronic communication and documents by some type of grouping, naming conventions, or categorizing. These grouping, naming conventions, or categorization may lead to other messages and documents that otherwise would have been overlooked or counsel might not immediately determine the association of this information.

Meta data analysis

Looking at the metadata of documents can assist in determining whether all locations of the documents were searched and provided. For example, looking at the location of the servers where emails were sent and performing pattern analysis could result in a conclusion that the recipient servers were not included in the document search. If an ECM system is utilized by the organization, identification of redundant storage systems throughout the network and/or review of whether the information is stored on off-line media becomes very important to ensure all anticipated ESI was actually produced.

Another example of analysis is a review of audit logs and history for information stored and/or managed in an ECM system, comparing them to the actual documents produced. As the ESI is always indexed (creation of some level of metadata) during the creation process, the forensic analyst should be able to perform meta data analysis related to whether information was deleted, produced, not produced, etc.

Content searches

Once the other steps have been taken, content analysis and searches can be performed to streamline tedious manual review of all of the information. At this stage of the process litigation support software and many eDiscovery applications are used to store the relevant information and enable counsel to being preparing the case. It is important to note that litigation software is neither an analytical tool nor enables full access to information stored in native format as each document is treated as a separate file thereby eliminating audit or historical data that is linked to the document or the communication threads amongst various individuals. The eDiscovery software has greatly matured over the past few years and provides tremendous value in preparing a case, but only after the analysis and technical examination has been completed. Otherwise, the question remains whether everything requested was received.

Conclusion

Applying these ESI forensic techniques can help litigators assist their clients in determining whether they have found all the materials responsive to a discovery order, as well, as review materials produced by opponents for completeness. As both an offensive and defensive tool, litigation strategies will be enhanced by knowing whether you got all the evidence covered by the discovery order.

Robert Blatt has more than two decades of experience helping clients with content management and workflow issues through the analysis, design and implementation of ECM systems. He is a recognized national and international subject matter expert in the ECM industry and is chairman of numerous national and international standards setting committees.

Virginia Jo Dunlap is a former securities regulator and litigator who developed processes along with Mr. Blatt to allow for review and analysis of large volumes of ESI in large-scale investigations and cases. In the private sector, she has served as a general counsel for a non-profit and as a senior executive in charge of risk assessment and mitigation for a global company, including finding practical solutions to ESI issues.