

Making Waves with Technology: An E-Discovery Case Study



Reprinted with permission from the June 2010 edition of *Legal Tech Newsletter*. © Copyright 2010. ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited.



Attorneys often become overwhelmed by the amounts of data involved in the discovery phase of a complex litigation case. I was recently involved in a case where one attorney accumulated approximately 445,000 documents for a litigated matter and had to respond to a Request for Production received from opposing counsel in 30 days. The attorney, who thought he could review every document himself, was drowning in documents and needed a life raft. The following case study demonstrates how attorneys can use technology to manage their discovery, meet deadlines and keep their heads above water.

Litigation Hold

The litigation involved a dispute over contract performance. Our client, the defendant, was a large government agency whose counsel's office was equivalent to a mid-sized firm and manages a significant caseload in multiple practice areas. Making several allegations in its complaint, the plaintiff raised over 40 separate issues. A litigation hold notice was sent to all potential witnesses and document custodians as soon as the litigation was anticipated. The attorney, however, did not actually begin retrieving documents and data from personnel until the complaint was received. All paper documents and electronically stored information were to be collected, processed, and stored in a repository in order to be reviewed and potentially produced in response to discovery. The attorney estimated that the data would be the equivalent of 80,000 pages of paper.

Document Retrieval

Once the attorney pulled the trigger on collecting the paper documents and electronic files, personnel were contacted. However, they were not always cooperative.

It is one thing for personnel to hold data and not destroy it, as that requires essentially no action. More effort is involved when personnel have to affirmatively gather documents and electronic files to turn over to a third party. This activity is in addition to their every day workload, which may already be quite burdensome. In this case, continual follow-up was required.

In order to achieve the most efficient retrieval process, point persons were established among the various groups of personnel involved in the contract and strict deadlines were set. Personnel created .PST files of their relevant e-mail that were burned to disks or copied to external hard drives for collection. In some instances, a specified location on a shared drive was created for personnel to copy their electronic data into self-named folders. This established one location to retrieve data from all of the custodians in a particular agency who were physically located across the U.S. in different office sites. Rather than having to go to each person's computer separately, all electronic files could be copied from one machine. Chain of custody was documented for all collected material. In addition, each document custodian was asked to sign a declaration stating that he or she had turned over all of their hard copy and electronically stored information to the attorney. This requirement provided the incentive for each custodian to be thorough so as to prevent personnel from finding more documents on the eve of discovery closing.

Point persons were also established for each of the large databases utilized by the parties during the performance of the contract. All electronic files pertaining to the contract at issue were collected from each database using an external hard drive. The

Making Waves with Technology: An E-Discovery Case Study

attorney made the decision to collect and hold the database files rather than process them. The files were most likely duplicative since the files loaded to each database would have been generated by custodians from that which was already collected. So the purpose of collecting the files was to ensure compliance with the court rules for data preservation as the databases were still actively being used.

The attorney threw the net far and wide in order to collect any and all material regarding the contract being litigated. As a result, approximately 445,000 documents (1.37 million pages) were collected from over 12 different agencies, 60 custodians, and four separate databases. This amount was over 10 times larger than what the attorney had first estimated. The deployment of technology would become the life raft.

De-Duplication

With multiple personnel working on the contract and utilizing the same documents, there was an abundance of duplicate files within the collection. Every recipient listed on an e-mail was turning over the same e-mail. Therefore, the first step in assisting the attorney with his discovery request was to attempt to reduce the quantity of electronic files collected.

For the more than 88 gigabytes of electronic data received, the MD5hash value for each file was used to identify duplicates. Analogous to a fingerprint, the MD5hash value is a quick and easy way to de-duplicate a population. Approximately 94,000 files were considered duplicative and removed. Processing costs were ultimately reduced since these duplicates were removed before loading the population to the document repository.

The next step was to assess what types of files were in the repository in order to identify common themes, such as photographs that wouldn't have text or files that don't process well because of multiple layers. A list of the various file extensions contained in the repository was reviewed. Seeing thousands of files with a particular file extension, it was determined that a multitude of drawing files had been collected. In this case, drawings were a specific request for production and highly unlikely to contain any privileged material.

The attorney agreed with the recommendation to have the files forego review and be produced in their native format. There were close to 9,000 of these drawing files produced to the opposition. This e-discovery decision saved the attorney time and money and reduced the review population further.

Text files were extracted from electronic files and scanned paper documents underwent Optical Character Recognition ("OCR") to create a text file. The majority of the hard copy documents collected were in good condition. However, any marginalia, handwritten notes or obscured text would have rendered less than optimal results in the OCR process. Based on the text generated from each document, duplication software was applied to the population in order to identify near duplicates. The technology is able to find any paper documents that are duplicative of electronic files, lesser-included e-mails or other similar documents that could potentially be a duplicate. It is important to note that any file that does not render substantive text, such as a photograph, would not be included in this process. By running the duplication software across the text files, it groups like documents based on the percentage of similarity that is set. In this instance, 80% likeness was used. The groups were then staged in a side-by-side review conducted by contracted attorneys. The result was the removal of 12,000 duplicates from the document population.

Auto-Review

Another process used in order to cull down the population further was auto-review. The largest-cost and most time-consuming step of discovery is the review phase. Therefore, deploying de-duplication technology and analyzing file types are worthwhile techniques to reduce the number of documents that ultimately need to be reviewed. An auto-review program will identify documents that have specific terms or phrases by searching the text files. When discovery is received, specific requests are made for particular documents and data. Auto-review is another tool that can be utilized to find the relevant and responsive documents you need to produce. You can also use an auto-review tool to flag privileged material by identifying documents with terms such as

Making Waves with Technology: An E-Discovery Case Study

“attorney-client privilege” or “attorney work product.” In this instance, terms were grouped in the following categories: relevance, attorney-client and attorney work product, proprietary, and Privacy Act. For each category, terms and phrases were developed in order to flag documents. Multiple variations of terms were considered. For example, the following terms would render these results:

- **Underwater motion device** – Documents with this exact phrase.
- **Fisher w/3 Paul** - Documents where Fisher is within three words of Paul whether Fisher falls before or after Paul, i.e. “Paul Fisher,” “Fisher, Paul,” “Paul J. Fisher.”
- **privileg*** – Documents with “privilege,” “privileges” and “privileged.”

Data analysts developed the scripts for the terms and their variations and ran them across the remaining document population. To ensure the terms were flagging the intended documents, sampling was conducted and the results were analyzed to determine false positives and misdetections. For example, the acronym of “ETC” standing for Everyday Tools Corporation would also flag every document with “etc” standing for etcetera. This term would render too many false positives in order to be considered valuable and be removed from the script. A misdetection occurred if the document contained a privilege term but was not, in fact, privileged. In this case, the percentage of misdetections was less than 1%.

By using the auto-review tool, the document pool was reduced by 45%. This provided more cost savings to the client in two ways. By using technology, the documents relevant to the litigation were able to be extracted and the document population for privilege review was now significantly less.

All documents with a relevant term hit were placed in a cue for immediate review. A two-step privilege review was conducted on the relevant documents: an initial privilege review and a quality check by a second more experienced reviewer. In this instance, contracted attorneys were hired to conduct the privilege review. In order to ensure the accuracy of the markings, two passes over each document was required.

Out of approximately 130,000 documents produced, there were 11 inadvertent disclosures (.008%). Opposing counsel, who also made some inadvertent disclosures, was cooperative and immediately destroyed its copies. However, if challenged, the two-step review process utilized would have most likely held up in court as the Federal Rule of Evidence 502 (b) requires “reasonable steps to prevent disclosure.” It is highly unlikely that a two-pass privilege review process would have been found not to be reasonable. See *Heriot v. Byrne*, 257 F.R.D. 645 (N.D. Ill. Mar.20, 2009), whereby the court held that the multi-step process used to produce documents entailed reasonable precautions to prevent disclosure.

Upon realizing that producing all of the responsive documents within 30 days of receiving the discovery request was highly unlikely, the attorney discussed a production schedule with opposing counsel. While the first production responding to the discovery request was made within the 30 days, rolling productions would continue over the next four months. The attorneys agreed upon .TIF images as the format for the documents.

Lessons Learned

Once litigation is anticipated, start gathering the electronically stored information. The size of a collection is unpredictable. Therefore, it is advantageous to conduct an early assessment to better prepare and plan your discovery management strategy. Once the complaint is filed, the data will be in one location and you will have an immediate estimate of the volume of data involved.

When developing the relevancy term list for use in an auto-review process, talk to the personnel from whom you received the documents and data. They are in the best position to know terms of art for the litigated matter outside of those terms used in the pleadings, such as acronyms or nicknames used in documents.

Recommend your client to participate in a meet and confer with opposing counsel. At this meeting, clawback agreements, term lists for auto-review processes, and production formats can be discussed. In this case, the attorney did not participate in a meet and confer and there was no clawback agreement in place at the time of the inadvertent disclosures.

Making Waves with Technology: An E-Discovery Case Study

If you have several stakeholders involved in the discovery process, such as in-house counsel, outside counsel, document custodians, and a discovery management vendor, form a steering committee with a representative from each group who has decision-making power. This committee will be able to discuss any issues that arise during the litigation and be able to render quick decisions to keep things rolling. It will also keep each group apprised of the other stakeholder's concerns.

Conclusion

With data accumulating daily in exponential proportions, the burden of managing data has become greater. Using technology to de-duplicate,

auto-review, and analyze your document population becomes a necessity. In this case, the attorney depended on his discovery management team to provide and implement the solution for meeting his discovery deadline. The attorney was rescued and able to focus on his litigation strategy rather than be in over his head in a sea of documents.

Betsey Pittard is a Project Manager at IE Discovery, Inc., overseeing the delivery of litigation support services to clients involved with contract disputes. She is a certified Project Management Professional and graduate of James Madison University and Widener University School of Law. She is licensed in Virginia, Pennsylvania and New Jersey.