

ELECTRONIC DISCOVERY & INFORMATION GOVERNANCE

Tip of the Month

**Making a Molehill Out of a Mountain: Tips for Handling Terabytes of Data****Scenario**

A medium-sized company is a defendant in a putative class action lawsuit. Outside counsel negotiated the scope of the plaintiffs' document requests as much as possible, agreed on a list of custodians, and sent an e-discovery vendor to the client's headquarters to collect email and copy hard drives. The vendor collects nearly a *terabyte* of data. The defendant's general counsel would like information regarding ways to manage the costs associated with this large amount of data, including considering various data analytics tools in the document review strategy and using skilled, experienced people who understand how to deploy these tools as part of a defensible process.

Goals

Key goals for companies that must respond to e-discovery requests are identifying the relevant data and identifying the data that is responsive to the plaintiffs' production requests. These categories overlap, but each allows the producing party to reduce the massive volume of data to a more manageable level. Using early case assessment tools and review workflow techniques, the case team may be able to prioritize the review and production. A benefit of prioritizing the review is that the most relevant documents are typically reviewed early in the process, which allows for early case assessment (ECA) and strategy development.

Choose Your E-Discovery Partner Well

When contemplating the use of advanced analytics for filtering large data volumes, hiring the appropriate e-discovery vendor is an important first step. Counsel should consider working with a qualified e-discovery vendor that:

- Can perform forensically sound data collections, process data using defensible workflows and prepare supporting documentation;
- Can make available ECA tools for filtering, searching and developing review strategies;
- Will host the results in a review application that facilitates further analysis; and
- Has the experience and the resources to support the case team and meet discovery deadlines.

In cases with a large volume of data, recurring "hosting charges" can become a real burden, especially during a long-running case. Vendors typically charge a per-gigabyte fee for hosting data. The use of various ECA tools can result in additional hosting charges. Counsel may want to explore negotiating alternative fee arrangements for processing and/or hosting at the outset.

Simple Steps To Take Before Responsiveness Review

In addition to prioritizing the review to move relevant documents to the front of the queue, some case teams may concurrently consider removing data that is unlikely to lead to responsive documents. The most common step is to “DeNIST” data during the initial processing in order to remove particular file types, primarily program or system files. Next, a case team may consider targeted searches to identify non-relevant files, often in the form of music, videos and photos. Similarly, “junk” email, such as daily newspaper reports and newsletters, might be culled prior to the application of search terms in order to minimize instances of false positive hits. By excluding these files, the case team might gain greater insight into the data while reducing the volume of data promoted to attorney review.

Another way to cull irrelevant material is through the use of date restrictions. By applying date filters, which often are agreed upon as part of the meet-and-confer dialogue with the requesting party, the case team can concentrate on a date-restricted set of documents for review and analysis. The case team might also consider custodian-specific time limitations. For example, if a custodian only worked in the relevant department for two months, there may be no reason to include email from that person’s entire tenure at the company. This initial cut can be performed during processing and excluded from the reviewable data.

Once broad cuts are made, the next step is typically to run search terms against the remaining data. Creating a list of search terms is an iterative process that is often developed through a process of discussions with the client and testing the terms against the database. The search term hit reports may suggest modifications of certain terms in order to identify relevant documents in addition to minimizing the amount of “false hits.”

Consider the Use of Data Analytics Tools

New technologies can make the review process more efficient and can get attorneys’ eyes on the key documents faster. For instance, “concept clustering” uses software to group emails about certain themes. Email threading can reduce review volume by showing reviewers only the most “complete” e-mail in a long chain and automatically coding its subsidiary parts so they do not need to be individually reviewed.

The case team might consider the use of predictive coding or technology-assisted review (TAR) tools during discovery and trial preparation. Although these tools were initially developed and marketed as a means for reducing the first-level attorney review costs, the focus today is trending toward using these tools to improve evaluation of both documents produced and those received in production. In addition, data analytics tools can be considered for prioritizing the review workflow; streamlining the second-level review, which is typically performed by outside counsel; and quality checking the review in order to prepare the documents for production. Data analytics can also save time, and potentially provide better results, during the preparation of witness files for depositions and trial.

Document, Document, Document

Whatever choices are made for data review, it is important to carefully document them. The use of these tools is relatively new and is still in the process of being fully understood by the legal community. As a result, a degree of skepticism can exist about the use of these tools. Thus, the case team is encouraged to work closely with their e-discovery provider to create supporting documentation that describes the process. This documentation can be used to replicate the process in future litigation and to explain and defend the process in the event of a challenge.

For inquiries related to this Tip of the Month, please contact Michael Frisch at mfrisch@mayerbrown.com, Patrick Garbe at pgarbe@mayerbrown.com or Kim Leffert at kleffert@mayerbrown.com.

Learn more about Mayer Brown's [Electronic Discovery & Records Management](#) practice or contact Eric Evans at eevans@mayerbrown.com, Ethan Hastert at ehastert@mayerbrown.com, Michael Lackey at mlackey@mayerbrown.com or Edmund Sautter at esautter@mayerbrown.com.

Please visit us at www.mayerbrown.com.