

eDiscovery Tips: Stranded Alone on a Desert Island with a Voluminous Document Collection?

An Overview of Tools to Make Your Task Manageable, Part 1 of 2

ChicagoLawyer

By Cathy Fetgatter and Lauren Allen

Reprinted by permission of Chicago Lawyer and LexTekReport.com. Copyright 2010 LexTekReport.com.



"A week of sweeping fogs has passed over and given me a strange sense of exile and desolation. I walk round the island nearly every day, yet I can see nothing anywhere but a mass of wet rock, a strip of surf, and then a tumult of waves."

- John Millington Synge

Discovery today can make you feel as though you are marooned on a desert island. The amount of potentially responsive electronic data has exploded in recent years, while litigation time frames and available resources have tightened. These trends have made discovery exponentially more expensive while organizations are facing unrelenting pressure on the bottom line. It's enough to make most attorneys feel lost and alone.

However, there are lifelines that can help you turn the data equivalent of a beach into a mere bucket of sand. With the right management and technological tools, you can streamline your e-discovery processes, saving time and money. By focusing particularly on the collection, de-duplication, search and review aspects of e-discovery, you can develop processes that are efficient, manageable, repeatable, and defensible.

It is important to note that there are no instant, out-of-the-box solutions. For every litigation, you must resist the urge to dive in—there may be shark-infested waters out there. Take the time to implement a thoughtful e-discovery plan. While standardization is a critical tool, no two situations are identical. Your methods may vary depending on the size of the litigation, your budget, the amount in controversy, and the willingness of opposing counsel to communicate. Therefore, establishing a structured flexibility to your e-discovery plan is paramount to your success.

In the first of this two-part series, we will explore the tools available for collection and de-duplication. Part 2 of this series will discuss tools you can use for the search and review phases.

Collection

Many cases go awry in the collection phase. By collecting too much data, you end up drowning in information and are forced to spend a tremendous amount of time and money to wade through it. By collecting too little, you are forced to recollect, which is redundant and expensive, or you risk missing critical documents that could jeopardize your case.

In order to reach the sweet spot between over- and under-collection, conduct ample research first so you begin with an accurate list of potential custodians. This is a living document that should be added to as new custodian information is discovered.

Once you are ready to contact the custodians, make the collection process as transparent as possible for them, eliminating any need for guesswork on their part. You should assume your custodians are not familiar with the document collection process. Be extremely specific about the data you want to collect, where this data may be housed, in what format it should be collected, and where the collection should be submitted. Do not assume the custodians have technical knowledge. Provide them with written instructions for conducting searches and capturing

eDiscovery Tips: Stranded Alone on a Desert Island with a Voluminous Document Collection?

the information to protect metadata. Have a template set of instructions already prepared that you can tailor according to the specifications of the case at hand. Document each touch with the custodians, starting with the litigation hold letter. Follow up any litigation hold notices at regular intervals. Your organization should have a policy on how frequently the reminders should be issued. If there is no such policy, this should be defined in a proactive discovery plan. You may consider having custodians sign declarations that they have complied with the collection. Declarations are particularly useful where you have recalcitrant document custodians.

Document the collection activities in a collection log. This tool manages the litigation hold notification process in real time, identifies problems and slow responders, demonstrates control over the document collection process and puts you in a position to quickly respond to questions or challenges from opposing counsel. These logs can also provide evidence of compliance to the court. You can use a database, spreadsheet, or other tool to create and manage these logs. Once you develop a standardized approach, you can save time and money while creating a repeatable and defensible document collection process.

Sometimes, you need to involve the IT department in your collection efforts. The IT staff knows where electronic data sits and you should consult them early. Frequently a shared, secure site such as an FTP site can be used for custodians to place all of their documents. IT can assist you with setting up such a site. IT can also help when you implement search terms. If a document custodian has left the organization, IT will be able to help access any files. When involving IT, the most important thing to do is to create a partnership and involve them in the process.

Once you complete your initial collection, it is key to step back and analyze what you have received. Are there any missing documents in the collection? Are you receiving a large amount of junk files or documents outside the relevant timeframe? Just as a hole can sink your lifeboat, a hole in your document collection can sink your case. Conversely, a document dump can be costly and sink your ship. At this point, you can analyze the scope of the discovery and open communications with opposing counsel.

De-duplication

Removing duplicate documents from the collection is another lifeline. The savings potential should not be underestimated. If a deduplication is performed before you dive into the review, you may drastically reduce data amounts thus decreasing review costs. In many cases, you can reduce the size of the initial collection by one-third simply by removing exact duplicates.

Many document collections also contain a large number of documents that do not reach the threshold of an exact duplicate, but meet the definition of a near-duplicate. By eliminating the appropriate near-duplicates, you can winnow your initial collection even further. Identifying near-duplicate documents is a sensitive process. Consider, for example, a Word document that has also been saved as a PDF file. The content will be exactly the same, but the two files would not be considered “exact duplicates” because they have different file hashes or document identifiers. When it comes to weeding out these types of near-dupes, the best approach often involves a software-based solution combined with eyes-on comparison.

Different tools offer different levels of deduplication sophistication. Some technologies allow you to set the threshold for identifying near-duplicate documents and then organize the matches in “like” groups of documents for your review. Less sophisticated methods, such as sorting by date or document title in Excel, can help identify potential near-duplicate documents as well.

Generally, it makes sense to de-duplicate for exact matches over the entire collection and by custodian. Removing near-duplicates should be decided on a case-by-case basis, depending on the goals of the litigation and the size of your budget, since it could involve a human review element.

When surrounded by a large document collection with limited resources to tackle it, there are collection and de-duplication tools that can successfully cull down your collection before the review even starts. We will explore tools that can be used in two other areas, search and review, in Part 2 of this series.

eDiscovery Tips: Stranded Alone on a Desert Island with a Voluminous Document Collection?

About the authors

Lauren Allen is a Program Manager with IE Discovery, a leading provider of Discovery Management services. She has been employed at IE Discovery since 2001 and Lauren is a licensed attorney with the Commonwealth of Virginia and a certified Project Management Professional. She can be reached at lallen@iediscovery.com.

Cathy Fetgatter has been a Project Manager with IE Discovery since 2008. Cathy has a J.D. and M.B.A from American University and is a licensed attorney with the Commonwealth of Virginia and the District of Columbia. Prior to IE Discovery, she engaged in private practice with an emphasis on complex commercial litigation. She can be reached at cfetgatter@iediscovery.com.