

Law Office Technology

Knowing When Advanced Electronic Discovery Technology Is Necessary

By Kevin Carr

The amount of discoverable ESI (electronically stored information) isn't simply increasing; it's multiplying. Some large corporations now generate literally billions of electronic documents each year. For professionals who have to field electronic discovery projects, it would be an impossible task — were it not for cutting-edge EDD (electronic data discovery) tools.

Certain cases, however, don't warrant all of the high tech. You don't need advanced discovery tools if basic, less costly systems can handle your task just as well. Before declining the latest offerings, though, it's important to understand what new EDD techniques, software and services are available.

What You Can Accomplish With the New Technology

The explosion of ESI and the impact of the newest Federal Rules of Civil Procedure (FRCP) have inspired an industry-wide emphasis on tools and best practices to help manage the data moun-

Carr is president of InterLegis Inc., a Dallas-based company that works with corporate legal departments, law firms and legal service providers to reduce the risk, complexity and cost of litigation, regulatory requirements and internal investigations.

tain in complex litigation. With upwards of millions of documents in even average size cases today, the ripple effect of not using advanced technologies can be significant in time and costs incurred. In a nutshell, today's more sophisticated tools inject a vital new phase into the litigation process: data mapping. With effective culling through data mapping, you confidently can eliminate the bulk of your document collection — the extraneous majority, sometimes more than 70 percent of documents — before it goes to the review team. And data mapping is becoming easier and faster.

Following are a few of the advanced technology options now available, should you require them. Some of these technology tools weren't available to litigation teams as recently as five years ago.

Data Clustering. Related documents are clustered to demonstrate the connections they share. New technology can sort the clusters, create sub-clusters and organize the entire clustered data collection. In addition to identifying relationships among documents, this also can eliminate huge blocks of unrelated documents.

Data Mapping. Data maps can chart the characteristics of data that must be considered by search and review teams. Before they ever begin full (and expensive) data processing, team members can look at the overall picture to see spikes in certain types of activity, run what-if scenarios and cull the document collection. Mapping the data prediscovery provides

a better understanding of the data as a whole, and helps to shape case strategy accordingly.

De-Duplication. Many documents — especially e-mail messages — are duplicates. Examples: carbon and blind carbon copies and memos sent to multiple recipients. Many other documents are near duplicates (an acknowledgement of receipt or a “thank you”). Good de-duplication software can find the original and cast out duplicates or near-duplicates with no substantial distinction from the original, primary document.

Visual Analytics. This more recent technology lets the review team see the big picture in their data collection through the use of graphical representations. They don't have to read all of the text to make connections; instead, they can immediately determine complex data relationships—document authors, creation dates, custodians who have handled the material, file format, etc., by viewing various formats of the data in graphical images.

Data Preservation. It's paramount that the original data collection be unspoiled. Modern electronic discovery systems go beyond that, providing snapshots of the document collection at each stage of prediscovery and discovery, from the initiation of the case to final disposition.

Other capabilities include pattern matching, similarity matching, e-mail thread and timeline analysis, and metadata indexing. An especially valuable EDD technique that has become common in recent years is relationship mining. Technology can identify characteristics of individual documents and, based on different document characteristics, find

similarities and relationships among items in the collection. And it can do this very quickly.

Attorneys must determine which 'state-of-the art' and 'cutting-edge' applications will best suit their needs for any given case. The question of when to use advanced tools versus more basic technologies is an issue that often gets lost when considering the best way to handle an EDD project.

Too Much Technology

Circumstances and particular elements of the case can help you decide which approach to take. There are two underlying questions: (1) How big is the document set? and (2) How complicated is the matter? Certainly, if your discoverable documents number only in the hundreds, your challenge is less daunting. If the case involves a single legal issue, exactly two opposing individuals, and no extraneous documents handled by other parties, your work could possibly be relatively simple. If — which today is more likely — the documents number in the hundreds of thousands or millions, if they've passed through the hands of many people, if they consist of a wide range of file types, or if they've been collected from many disparate sources, you almost certainly will benefit from advanced technology.

There are less obvious signals than the sheer size of the data collection, the number of people involved or data type that might influence the tools chosen for a given matter. Before you decide whether EDD power technology is overkill, you need to assess the features of your case very carefully.

Determining Factors May Include:

The complexity of the legal issues. Are they simple and straightforward, or intricate? Are they with or without clear legal precedent? Are multiple issues involved, either separate or interrelated? Will the documents likely be needed for litigation in the future?

The number of authors and recipients whose documents must be dealt with.

Does a legal issue hinge on a modest-sized document set that originated with a single individual? Might the actions of others, based on the contents of a document set, have a bearing on a disputed matter?

The number of custodians who handled the data. How many hands did different pieces of the document collection pass through? How many computers? Are the custodians in multiple locations? Data collected from only a few individuals in a single location is very different from a case that includes data from several dozen or even hundreds of custodians across several locations, departments, divisions, subsidiaries — or even countries.

The diversity of file storage formats. If some of the documents aren't in a reviewable format, your EDD team has a problem. Most e-mail is stored in Microsoft Outlook format (.PST), but you may encounter exceptions, particularly with documents that were sent via a free Webmail service.

The extent to which e-mail files come into play may have a significant impact on the level of technology required in any given case. E-mail is very dynamic, and most discovery data collections today include vast amounts of e-mail. Basic technologies may only process the most basic bibliographic data associated with e-mail files, such as authors, recipients and dates. But in most cases there is greater context within the files that simply cannot be ignored. E-mail files contain complex threading, embedded files, multiple recipients — some carbon copied, some blind copied, and countless duplicates and near-duplicate versions of many files. As such, oftentimes a dataset containing only 10 gigabytes comprised entirely of e-mail requires more advanced technology than a dataset containing 50 gigabytes of Microsoft Word files, for example.

Another example that demonstrates a distinction between the need for advanced versus more basic technologies is the number of file types included in the data collection. There are more than 2,000 commonly used file types known today; approximately 200 of these have the most widespread use, including Word, Excel, Lotus, and WordPerfect files. When a medium-size or even a large dataset is

limited to a single file type, such as simple word processing files, basic culling and review technologies will likely suffice. Fancy bells and whistles might still be nice to have, but aren't necessarily critical to the success of the case. On the other hand, if the dataset includes several hundred various file types, each with its own unique structure of metadata, a more advanced tool will quickly pay for itself.

The complexity of a matter and of the data itself also plays an important factor in determining the type of technology required for the case. For example, the litigation involved in a two-vehicle automobile accident on a quiet country road caused entirely by driver error is likely to be far less complex than an aviation litigation case. The latter may include more than 200 passengers that were killed, and faulty mechanics manufactured by a company other than the aircraft manufacturer. Airlines, airport commissions, air traffic controllers, airplane manufacturers, passengers, victims' families and many others might be actively engaged in a case of this nature for years. And in this type of litigation, the technology required to successfully navigate through several stages of the discovery life cycle is tremendously more sophisticated than that which is required in a less complex two-vehicle car accident.

Know Your Options

Your objective is to manage ESI as effectively as possible. You want to be able to devote most of your resources — staff and budget — to case planning, not to data analysis. Not every case requires the most advanced technology available today, depending on the nature of the case and especially the size and scope of the data involved. However, in complex cases the newest EDD systems can pay for themselves many times over by culling irrelevant documents, pinpointing those you really need, and managing your dataset effectively. In doing so, they will help you win the best possible outcome for your client. In today's litigation landscape, it has become essential for legal professionals to understand what technical assets are available and when to apply them. ■