Pricing Processing in E-Discovery: Keep the Invoice from Being a Surprise

By Seth Eichenholtz

E-discovery continues to confound law firms and clients with a lack of consistency in pricing. This lack of consistency may have many reasons, likely due to various vendors having different internal costs owing to the wide range of core competencies and services that e-discovery vendors purport to offer. But perhaps the most critical reason for the lack of consistency in pricing is that vendors have been able to take advantage of the fact that both law firms and clients are all over the map in defining specifically what "processing" means in the e-discovery lexicon. Processing data in e-discovery means different things to different people, and pricing for processing has followed suit.

Of all the major hurdles in any e-discovery engagement, perhaps the biggest challenge is being able to foresee the costs of processing the data. Processing can also become the most expensive component of litigation, short of a long and drawn-out document review. The reason for this is that there has been no uniformity among e-discovery vendors regarding how to categorize the various elements and issues inherent in processing data in e-discovery nor what those costs should be. And it is very easy to envision a case with current market processing costs where the processing costs alone can quickly reach six-figure sums. So to the extent that one can police, lower, and perhaps best of all, predict pricing, this is a critical issue and an opportunity to better represent your company or client.

Before we assess pricing processing, we should first define what processing in e-discovery means. While there are no governing bodies overseeing e-discovery, there are a few key groups and organizations that are well recognized and respected for their insights and descriptions of the various components of e-discovery. The E-discovery Reference Model (EDRM) (http://edrm.net/) and the Sedona Conference (www.thesedonaconference.org) both offer well-documented, court-cited definitions of what processing in e-discovery is.

The EDRM states the following about the processing stage (lifecycle) of e-discovery:

> [F]ollowing preservation, identification, and collection, it often becomes necessary to "process" data before it can be moved to the next steps of the lifecycle. Some primary goals of processing are to discern at an item-level exactly what data is contained in the universe submitted; to record all item-level metadata as it existed prior to processing; and to enable defensible reduction of data by "selecting" only appropriate items to move forward to review. All of this must happen with strict adherence to process auditing; quality control; analysis and validation; and chain of custody considerations.

The EDRM goes on to break up processing into four subcategories: assessment, preparation, selection, and output. Assessment may allow for a determination that certain data need not move forward. This stage, typical in a consultative manner, may include general data sets or evidence items that are deemed not relevant or necessary for whatever reason. The preparation phase involves performing activities on the data set that will later allow for specific item-level selection to occur (extraction, indexing, hashing, etc.). The selection phase involves de-duplication, searching, and analytical methods for choosing specific items that will be moved forward. The output phase allows for transport of reviewable items to the next phases of the lifecycle.

The Sedona Conference includes the following regarding the processing of data:

> Principle 11: A responding party may satisfy its good faith obligation to preserve and produce relevant electronically stored information by using electronic tools and processes, such as data sampling, searching, or the use of selection criteria, to identify data reasonably likely to contain relevant information.

While both the EDRM and the Sedona Conference offer excellent points on methodology, there is little information in either definition to truly define processing as it would appear in a vendor's statement of work or services agreement.

However processing is defined, the following key steps are prevalent:

- de-duplication (either globally or by custodian)
- filters for relevant dates, custodians, file types (de-NISTing, or removing known files that have no pertinent data in them, like system executable files)
- search terms (or keyword searches, also called Boolean searches)
- metadata filters
- load file creation

The trend seems to be employing all of the steps listed above before the load file creation stage—often called the initial culling phase—and price it on the lower end of the spectrum. This is most likely a knee-jerk reaction to the general complaint by law firms and clients that e-discovery costs too much. But

Seth Eichenholtz is vice president, e-discovery manager with Swiss Re, based out of Armonk, New York.

Published in *Pretrial Practice and Discovery*, Volume 19, Number 2, Winter/Spring 2011.

this brings up two important questions:

- When was the last time you purchased anything of importance because it was the least expensive?
- If vendors drop their price, does that mean they are making less money or that other parts of their services are being priced appropriately?

The first question is really a warning to not be seduced by the lowest price. Many vendors choose to charge a low upfront cost to cull down the initial data set as an inducement for the client to sign on. The second question is a logical follow-up to the first, postulating that a low price on one part of the bill may be making up for various other components. This is what is going on all too often in e-discovery processing.

The initial culling involved in processing seems cut and dry, but there are issues to consider that many vendors will not list in their pricing sheets or are purposely keeping vague. The most common issue I've seen in the initial culling phase relates to keyword searches. As anyone reading this article knows, generating searches and keywords is far from a perfect science. And I've rarely been involved in any matter where the search terms are set and not altered (or more likely added to) throughout the engagement. How your e-discovery vendor charges for that service is critical from a cost perspective.

Let's see how this plays out with three different pricing structures.

- Vendor A includes keyword searches (and ostensibly, search terms as well) but no further information in its pricing.
- Vendor B includes initial keyword searches as part of the initial culling price and additional keyword searching at an hourly rate.
- Vendor C includes keyword searching as part of its initial culling stage and up to three iterations before load file creation.

It is entirely possible, and arguably even likely, that the first set of keywords or search terms will be modified. Vendor A has not provided insight on how they would charge for three or four different iterations of search terms. Ostensibly, it could very well charge the same per-GB charge for initial culling for every new set of terms. While this method may inspire counsel to come up with a complete set of search terms to avoid unexpected costs, the client will most likely be unhappy with re-processing data over and over. By applying some numbers to this example, one can see that at a low market rate of $150 per gigabyte for initial culling, if the data set is 50 gigabytes and the search terms are re-run four different times, the client may think it is paying $7,500 ($150 multiplied by 50 gigabytes) for the initial culling, but without clear language or discussion on pricing, the client could receive a bill for $30,000 ($7,500 multiplied by 4).

Vendor B spells out clearly that after the first set of search terms, it charges an hourly rate to re-run search terms. This is still an incredibly vague way of assessing costs, however, since there is no way to predict how long it would take to re-run search terms across a given data set. Variables like software type, type of data, volume of data, and processing power are but a select few variables to consider. But it is safe to assume that even in a modestly sized data set of 50 gigabytes, it is a matter of hours, not minutes or seconds, to re-run search terms. In applying some numbers to this example, assuming the same $150 per gigabyte initial culling cost, the client needing to re-run search terms four different times, a low market rate of $50 per hour to re-run search terms, and assuming a conservative estimate of four hours to re-run each new set of search terms, the client would go from paying $7,500 (the first round of search terms included) to $8,100, or an additional $600 for running search terms ($50 per hour multiplied by four hours, or $200 performed three times).

Vendor C allows three different iterations before load file creation for its clients to finalize their search terms. To some, that may be fair and practical, but to others, it may not mean a thing if the matter is a multi-district, multi-party litigation with nine different law firms involved where running three sets of search terms is a drop in the bucket. Ostensibly, Vendor C can charge its initial culling fee after any new search term is needed beyond the first three. Applying numbers to this example is not necessary to clearly show that in using a vendor who prices this way, one must have a solid grasp of the keyword searching that will be needed for that specific matter. If it will clearly go beyond three rounds, a separate pricing agreement should be made to ensure that the initial culling costs don't skyrocket.

Unfortunately, there is no best situation here, as each case and situation is always unique. The key is to be able to look at a service agreement or scope of work and be armed with the knowledge to extrapolate how things may play out so that the client is not stuck with a bill that is thousands of dollars higher than what was estimated.

The critical point here, and the best way to be a true advocate for your client or company, is to get the most consistent pricing—not necessarily the least expensive. If you estimate a certain amount for e-discovery costs, coming as close to that estimate is much more likely to be beneficial than having the cost be the lowest you could find with the risk that the service will suffer.

Load Files

This is an area that tends to be so consistently mismanaged that it is interesting more people are not discussing it. First, some context and a definition for the uninitiated: E-discovery projects frequently involve attorneys reviewing documents in TIFF or PDF images. Converting native files to TIFF or PDF format is frequently standard operating procedure for many attorneys and is a common method for producing documents to opposing counsel. But creating TIFF or PDF images also removes the metadata from the original document. Therefore, a separate file containing each document's metadata (typically in the form of a spreadsheet with specifically defined fields) is needed to review the data or produce it. The spreadsheet is the load file. Some vendors call this step deployment for review, stage two processing, or native file processing—and it is almost always more expensive than initial culling due to more manual labor.

The EDRM offers a good general definition for a load file: "A data file that sets out links between the records in a database and the document image files to which each record pertains."

(*See* http://edrm.net/resources/glossary/l/load-file.) The key, for e-discovery purposes, is that the load file contains all the metadata from the original data set, points to the extracted text for search functionality in a text file, and links the metadata and text files back to the image files that were created.

For document review, a load file containing a production set is typically created by an e-discovery vendor to easily input information into a document review repository tool like Summation, Concordance, or iPro. Things can get quite complicated, however, as load files come in a variety of formats and versions. And now with law firms, clients, and vendors all using different tools for different purposes, you may find yourself in a situation where data is sitting in one tool and a load file needs to be created to move the data to another tool for document review, but if the load file specs don't match up or the vendor does not have a tool readily available to create the necessary load file, extra costs for creating that custom load file will certainly be incurred.

Thus, to keep costs in check, it is vital to understand what document review tools are going to be used by both parties (and any additional software, like Early Case Assessment tools) as well as what fields will be needed in your document review as early as possible, as the corresponding load file specifications will be a determining factor in the cost.

Many law firms and some clients have very specific, and oftentimes extensive, specifications for how data is to be processed and how metadata and text are to be extracted. While there are common load file formats, it is absolutely critical to understand how your vendor is charging for the creation of the load file. As mentioned earlier, from the vendor's point of view, this is a different procedure than initial culling with different software and additional time required. It is the critical link between the raw processed data and the processed data that the attorneys are reviewing.

The load file processing stage can yield the following components:

- extraction of metadata
- extraction of text (and creation of corresponding text file)
- native files
- creating image files (TIFF or PDF)

Most vendors will include the extraction of metadata and text, and of course, the corresponding native files in their load file pricing. The pricing discrepancy lies in how vendors charge for the creation of image files. There are effectively three ways to deal with image files at this stage:

1. Create the images during the load file creation stage before the data is put in the review tool. This is the most desirable approach, particularly if there is a need to search the corresponding image files. This approach also has benefits when production timelines are tight because the images are already created.
2. Use the TIFF "on the fly" during document review approach when small or unanticipated productions need to take place. This may create issues with bates numbers and branding those image files with and among various users.
3. Create the images after document review while gearing up for production. This approach is very common assuming there are no tight production deadlines.

When you create the TIFF and PDF image files can also be an expensive decision. Since vendors typically charge $.02, $.03, $.05 per page or more to create a TIFF or PDF image while reviewing the documents, one must consider if it is worthwhile to create the images early on during the load file creation stage, or during or after the document review at a per image price. What I've seen repeatedly is the client choosing the less-expensive option on first glance, which is creating the TIFF images after the document review, and then being baffled as to why there are additional fees in the tens of thousands of dollars for the creation of those images.

As an example, let's assume your data set is about 50 GBs. If you choose to create the files during or after the documents are reviewed, let's assume only a percentage of the files will be needed in TIFF format, bringing the volume of documents down to 25 GBs. Assuming 50,000 pages in a GB at a conservative rate of $.03 per page, it could be an additional $37,500 in costs—most likely not in the original estimate.

Now compare this to the situation of choosing to create the files during the load file creation stage with the same assumptions. If the additional costs accrued were $700 per GB to the price of the load file creation cost (a market-accurate number), even without the hindsight of only converting half the volume, 50 GB multiplied by $700 is $35,000. Not only is that actually less than if half the documents had a TIFF image created during the document review stage, but the cost would

Practice Tip for Young In-House Lawyers

Make yourself accessible to all members of the organization; make it a priority to get to know not only upper management but all other levels of the organization's employees. These people can be as much of a resource to you concerning institutional history and operations as you are to them regarding legal matters. In addition, the more people who are familiar with you and feel that they have access to you, the greater the likelihood that issues—both legal and otherwise—will be brought to your attention. Even if the issues are not legal issues that are appropriate for in-house counsel to address, they can be passed on to the appropriate administrator. Join a listserve in your practice area so that you can rely on the wisdom and advice of others who have more and/or different experiences in the same field.

—Jean Walker Tucker, Senior University Attorney, University of South Alabama

Published in *Pretrial Practice and Discovery*, Volume 19, Number 2, Winter/Spring 2011.

have been a known entity and accounted for. It is exactly this kind of pricing challenge that has driven lawyers and clients to become weary when working with e-discovery vendors.

Be sure to consider if image files will be utilized early on and query your vendor regarding what services are included in the load file creation phase. Some vendors will offer all-inclusive pricing on the load file creation, which may appear to be more expensive, but most likely will keep costs down and, moreover, becomes a quantifiable cost early on. Other vendors may offer a blended rate that includes the initial culling phase and the load file creation phase. Either way, be sure to understand and account for all the potential costs along the way. The goal is to never be surprised by the invoice at the end of the day.

Other Costs to Consider

Project Management Fees

This may also be identified as consulting or professional services fees. Many vendors will attach project management fees to your processing costs. To some extent, this is understandable, as there are certain things that must be done in the processing phase that take up time, such as filling out and reviewing for quality control chain of custody forms, generating reports, and communicating with the client and/or law firm. Some of this may be included if your vendor offers flat-rate pricing, but be sure to ask, and never assume anything when reviewing a services agreement. Pricing varies on this greatly, from under $50 per hour to upwards of $275 per hour. Those numbers can add up very quickly and are an expensive afterthought.

Technical Time

Like anything with computers, there is the very real concern that there will be technical issues to overcome. From the mundane (password-protected files) to the complicated ("My data is in Early Case Assessment Tool X, and I need a load file created with custom fields for Review Platform Y by tomorrow afternoon") most vendors will charge you some form of an hourly rate. Be sure to query your vendor as to what constitutes actual billable technical time as opposed to something that should be covered under the general processing fee.

Exceptions Processing

In any data set, there will most likely be files that the vendor was unable to process upon initial culling. Most vendors will supply their client with an exceptions report listing those specific files and related information. These reports will typically include odd or proprietary file names/extensions, container files (like a ZIP or PST file—not to be confused with the actual files contained within), password-protected files, and others. The exceptions report can be critical, and the pricing methodology varies among vendors greatly. Some vendors will include processing the contents of the exceptions report as part of the initial culling phase; some charge extra for that kind of work (hourly or by the GB); and many will combine a per-gigabyte pricing fee with an hourly fee for technical time. Be sure to confirm how your vendor charges for dealing with exceptions reports before the initial culling begins.

Conclusion

This article is not intended to give in-depth instruction on how to process data nor is it intended to give ammunition to law firms and clients to be overly aggressive with their vendors. The fact is that most vendors really try to do a good job in servicing their clients. But e-discovery presents many unique challenges, and, combined with the wide variety of backgrounds vendors come from, the lack of any true governing body overseeing e-discovery methodology, and the inherent miscommunication between attorneys, IT professionals, clients, and litigation-support professionals, there are myriad issues and challenges that arise daily. However, by grasping the basic nature of what the processing stage in e-discovery truly consists of, and, moreover, the nuances of how various vendors price those services, one can learn to ask the right questions and thus engage the best vendors for the job. The best possible result is not only a job well done but also a price tag consistent with expectations and delivered services.

"Pricing Processing in E-Discovery: Keep the Invoice from Being a Surprise" by Seth Eichenholtz, 2010, In-House Litigator *25:1, pp. 1, 18–21. Copyright © 2010 by the American Bar Association. Reprinted with permission.*

The Ten Commandments

(Continued from page 8)

a written contract. By submitting cross-motions for summary judgment, the parties are telling the court that they do not think that there are any genuine disputes as to any material fact and that the case should be decided as a matter of law. The court, however, does not have to agree with that assessment, but is free to conclude that there is a disputed question of fact and therefore deny both of the cross-motions. *Podberesky v. Kirwan*, 38 F.3d 147, 156 (4th Cir. 1994), amended on denial of reh'g, 46 F.3d 5 (4th Cir. 1994).

Conclusion

Motions for summary judgment are powerful weapons that are frequently used in contemporary civil litigation. Careful attention to the applicable rule provisions, together with knowledge of the substantive legal principles governing your dispute, provides the necessary foundation for analyzing the availability of summary judgment in your case. If you build on that foundation by following the foregoing "Ten Commandments," you should do well in representing your clients' interests.

Published in *Pretrial Practice and Discovery*, Volume 19, Number 2, Winter/Spring 2011.