

# BAR BULLETIN

**KCBA** KING COUNTY BAR  
ASSOCIATION  
Justice... Professionalism... Service... Since 1886

This is a reprint from the King County Bar Association Bar Bulletin  
May 2016

## A Better Way: Document Mining vs. Document “Review”

**By Larry G. Johnson**

“Document review” is the Dante’s *Inferno* of civil litigation. Since almost 100 percent of documents these days are digital or have a digital origin with embedded useful metadata making them preferable to printouts, such “reviews” are conducted on computers, usually lots of them all lined up in a big room.

People who do this work are often found in front of regimented rows of monitors set up in the cavernous bowels of the building housing the law firm way upstairs, or in a distant warehouse. The reviewers typically consist of minimally paid contract lawyers or paralegals, all working on the clock. These austere venues can have a Silicon Valley feel to them, complete with discarded pizza boxes on the cheap carpet and sleeping bags here and there for the all-nighters.

Such a room comprises the set in an episode of AMC’s “Better Call Saul” series, where the girlfriend of the con-man lawyer Saul Goodman finds herself banished to the law firm’s dungeon-like doc review basement as punishment for seemingly involving herself in one of Saul’s frauds. These doc review rooms can be as dark, grim and quiet as any sullen sweatshop in a third-world country, with the difference being that something of beneficial use can come out of sweatshops.

### Is Your Client Getting His/Her/Its Money’s Worth?

In a case I was involved in not so long ago, there were hundreds of thousands of documents that were distilled after extensive review to a manageable

few hundred for use at trial. I calculated the average cost for *each* such document making the final cut at around \$3,300. So many haystacks, so few needles.

It is widely understood that by far the most expensive part of the e-discovery process is the component of the human document review of usually thousands of computer-generated documents and emails to cull only the non-privileged, relevant ones for production to a requesting party.

Sadly, though, this customary review process is incredibly inefficient and unduly time consuming, and hence clients often pay a very hefty price that is hard to justify based on the results obtained.

### It Doesn’t Have To Be This Way

Given the many pressures in preparing a civil case, wouldn’t you rather be free of doc reviews altogether or dispense with most of the process, since so much of it anyway is based on cynical gamesmanship and only incidentally part of a search for the truth? Is there a better way to get useful information rather than doing the death march involving many lawyers and paralegals staring at one document after another on computer screens until their brains go numb?

Of course, “doc review” can be variously defined as perceived by different lawyers. On the one extreme is the position taken by a major Wall Street Big Firm as explained to me across a conference table that a helicopter could land on: “Our associates here do not do ‘document reviews’ in our litigated matters,” this lawyer sniffed. “Their job

is to learn intimately the businesses of our clients by reading every single document, one by one.” I thought I heard the distinct ringing of a cash register as he ended that sentence. I had to wonder what it must cost for all these \$700/hr. associates tutoring themselves randomly by reading mountains of tedious old documents with no particular relevance to anything.

At the other extreme I suppose, if you want to call it that, is me: Give me access to the contents of a person’s smart-phone, and I’ll probably have most of what I need to know from him or her in the matter: saved and deleted text messages (more unguarded in content and exchanged in greater profusion than emails); saved and deleted voice mails; saved and deleted emails; GPS data in photos taken on trips; credit card and other financial information; and a portal to all the user’s social media.<sup>1</sup>

*But*, you ask: “If I have to respond to production requests for all my client’s electronically stored information regarding the issues in a case, how can I know what documents are responsive if I don’t look at them?” Well, fair enough, but think about it — literally speaking, do *you* in fact *look* at all that ESI? Probably not.

Instead, your firm or outside e-discovery vendor/consultant employs various software tools to search text with search terms, often in various logical “Boolean” combinations to exclude as much chaff in the data universe of assembled ESI as possible in order to reduce the size of the ultimate document review set. Then the junior associates,

---

contract attorneys or paralegals do the drudge work of skimming through and categorizing the hopefully manageable bunch of documents for the final cut — not you.

But I submit that it is precisely in that stage of customary document review arrangement where the e-discovery process breaks down and becomes unjustifiably inefficient and expensive. By that point the doc reviewers are reduced to using a limited technology platform to glance through one doc after another, skimming through most of them, and now and then tagging some for relevance, perhaps including the relevant issue(s) to which the document relates.

That process harbors an Achilles heel: At this point in the e-discovery process, the review has been removed from the source machine(s) whence the info came, so the reviewers are divorced from the original computer context and can't go back to look at it directly. More importantly, most if not all of the selected reviewers have no trial experience, they don't have the litigator's instinct for what makes for a red flag in a document or a red herring, or how a document might expose a vital clue or anomaly in the overall chronology or issues landscape of the case.

Under current doc review practices, what you get instead is some pretty thin gruel: 1) You can say you exhaustively did your due diligence to look for and produce relevant information under your Rule 26(g) certification obligations, and 2) You have the time logs and huge expenses incurred to prove point No. 1. In other words, you do have your props in place if needed to defend against the usual kinds of faux outrage in the opposition's motions to compel, claims of spoliation, discovery abuse, and all the rest of the usual, tired gamesmanship. But in terms of evidentiary substance, what do you really have?

### **Have You Done a Worthwhile and Ethical Job for Your Client?**

Here is an alternative universe for what would probably be more effective ESI searches and productions of ESI, both in terms of costs and results: data *mining* vs. "reviewing."

In my KCBA Bar Bulletin article in

the February 2016 issue,<sup>2</sup> I postulated that the brain and instincts of an attorney experienced in trying lawsuits and in e-discovery technologies is the kind of "Pro from Dover"\* you need in order to get quickly to the heart of a case and win it. You can hire or associate such an "e-discovery lawyer" to handle all aspects of ESI discovery in your case and present the final results from his/her efforts to you, whereby the overwhelming bulk of the work will have been done for you (and independently CR 26(g)-certified). Your job would be chiefly limited to verifying and approving the pro's work product leading to the ultimate production to the other side.

That is probably pretty much what you do already in your current document review setups, anyway, but the results will be improved from much higher quality inputs.

### **What Special Skills Can Such a Pro Bring To Streamline and Improve the ESI Production Process?**

Here are such skills in a nutshell:

1. Computers already have several built-in, self-organizing features. For example, some folders for files of a certain type come already installed on the computer, like "Music" or "My Documents," and there's a "Recycle Bin" where deleted documents go. Several other examples could be cited.

Most users learn quickly to set up their own unique file folders for saving documents, and they usually give the folders content-related names, such as "2016 tax documents." Moreover, files saved, however and wherever saved on the computer, are automatically date- and time-stamped by the computer in ordinary computer usage, along with other useful metadata. Email metadata, in addition, can tell you who sent them and via what computer and network servers, and the email message adds context to the importance and intent of attached documents.

A savvy pro knows where to look for key information on a "live" computer (or cloned version of it) after first safely making and preserving a cloned copy of the computer's hard drive(s) to avoid potential spoliation claims. Much useful information can be derived by

looking at a computer the way the user saw and used it. It's like the benefit to be gotten from a doctor's examining a live person to appraise the state of his health rather than trying to imagine how that person's health must have been like when alive by examining his cadaver.

2. The use of search terms to find relevant ESI is as much an art as it is a science, maybe even more so. Of course, specialized skills count. There are various kinds of search syntax options one can employ (e.g., Boolean, proximity and fuzzy logic searches) that can help to refine text searches so that narrowly focused results are more likely, and it takes skill to formulate those; but of equal importance is deciding what sets of data to focus on (Just the key witnesses? Only the CEO? Just the Facebook entries? All of these at the same time?).

Also critical in a "live" or "live-simulated" examination is: How do you take advantage of the first responsive documents you find so as to let them inform subsequent text searches? Are the initial results clues to use as new search terms? Is there a helpful buzzword, such as a project code name or a person's nickname, that you can now take advantage of as a search term?

This kind of fluid data navigation and data mining is what can happen when you take an iterative and dynamic approach, much like a pilot flying through a storm plying the stick, flaps and rudder, responding to what transpires in front of him or her to reach the goal of landing safely.

3. Use existing computer forensic software and hardware tools that are anchored in probabilistic modeling and math, so that search result document sets can be justifiably reduced according to relevance ranking based on algorithms that tell one how well document "hits" are semantically related to your search term(s). My experience and that of others show that the time — and hence also the cost — of finding and isolating relevant documents is typically drastically reduced, by as much as 80 percent.

The probabilistic model approach also comports well with the new emphasis on proportionality as embodied in recent changes to FRCP 26 and equiva-

---

lent state rules.

The “Pro from Dover”\* who architects protocols along these proposed guidelines will be an ideal independent expert to testify, if needed, to the accepted best practices and defensibility of the approaches as outlined.

### **Time To Change**

If you don’t currently see for yourself the many inefficiencies and unnecessarily high costs in the way doc reviews are usually conducted by law firms in cases of even modest complex-

ity, expect eventually to see your institutional clients point them out for you.

---

\*I will give a \$10 Starbucks gift card to the first person who can correctly guess the title of the movie where this expression was coined, though you definitely risk dating yourself if you claim the prize. ■

---

*Larry G. Johnson is a lawyer in Newcastle, and has been a member of the Washington bar since 1974. He recently served on the E-Discovery Subcommittee of the WSBA Escalating*

*Cost of Civil Litigation (ECCL) Task Force. Besides being a litigator, for the past 20 years he has served as a consultant and expert witness in e-discovery matters. He does business as Electronic Data Evidence ([www.e-dataevidence.com](http://www.e-dataevidence.com)).*

---

<sup>1</sup>For more on data-rich smartphones, see my article in the October 2015 KCBA Bar Bulletin, “Smartphones: A New ‘Safe Harbor’ for E-Discovery,” at [http://www.e-dataevidence.com/2015\\_10\\_Reprint\\_Smartphones.pdf](http://www.e-dataevidence.com/2015_10_Reprint_Smartphones.pdf).

<sup>2</sup>“Bits, Bytes and Brains: Still the “No-Brainer” E-Discovery Tool.” [http://www.e-dataevidence.com/2016\\_01%20brains%20as%20E-Discovery-Tool.pdf](http://www.e-dataevidence.com/2016_01%20brains%20as%20E-Discovery-Tool.pdf).