BOTTOM LINE DRIVEN PROPORTIONAL REVIEW

Ralph C. Losey
Jackson Lewis, LLP

I have been focusing on the problem of high e-discovery costs since 2006. At that time I phased out my general trial practice and devoted all of my professional time to electronic discovery law. I focused on the expenses associated with those legal services because it was obvious in 2006 that unpredictable, high e-discovery costs were a core problem in civil litigation. They still are. Uncertain costs keep many attorneys away from e-discovery, even though nearly all original records and documents are electronic, and have been for many years.

The price and uncertainty of e-discovery costs are still the main reasons attorneys try to avoid e-discovery, even though nearly all original records and documents are electronic, and have been for many years.

The most expensive part of e-discovery is the document search and review process. In most projects it constitutes between 60% to 80% of the total cost. *See Where The Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery* (corporate survey found an average of 73%).[1]

Ralph C. Losey is a partner of Jackson Lewis, LLP, a national labor and employment law firm with 775 lawyers in 52 offices. He leads the firm's Electronic Discovery practice group and limits his practice to e-discovery issues. Ralph can be reached at Ralph.Losey@JacksonLewis.com. He is the author of five books on electronic discovery published by West Thomson and the ABA and e-Discovery Team, and is the publisher and principal author of the *e-Discovery Team Blog* at http://e-discoveryteam.com. Ralph has limited his legal practice to electronic discovery since 2006, with a special interest in the last few years in the use of *artificial intelligence* to find and review evidence in *big data*. Ralph has also served as an Adjunct Professor of Law at the University of Florida teaching introductory and advanced electronic discovery. Ralph has been involved with computers and the law since he began private practice in 1980. His full biography may be found at RalphLosey.com. The opinions expressed in this article are his own, and not necessarily those of his law firm, clients, or University.

[1] Report can be found online at http://www.rand.org/pubs/monographs/MG1208.html

For that reason, although I practice in all areas of electronic discovery law, my efforts to control costs have focused on document review.[2]

Two Different Review Tasks

The review costs in turn arise primarily from two legal activities:[3]

1. Search for and identification of the likely responsive or relevant documents. (For convenience and simplicity purposes this paper will refer only to *relevance* henceforth, and not responsiveness, although the author recognizes there can be differences.) This is a binary decision, *yes or no* as to relevance. This type of review is often called *first pass* review.[4]
2. Study of the documents identified as likely relevant to determine which must be withheld, logged, redacted, and/or labeled to protect a client's confidential information, such as privileged communications. The subsequent reviews can also include specific issue tagging work unrelated to confidentiality concerns. Documents not identified as relevant are not included in these further reviews, but may be subject to sampling for quality assurance purposes.[5]

The second reviews for client confidentiality purposes can be very problematic, expensive, and risk-filled.[6] For instance in *Tampa Bay Water v. HDR*

---

[2] There are essentially ten tasks that lawyers perform in e-discovery, three of which pertain to document review. They are described in the *Electronic Discovery Best Practices* found at EDBP.com. The three types of services pertaining to document review can be found at: http://www.edbp.com/search-review/. The functions performed by electronic discovery vendors, which includes non-legal services such as data processing, are outlined in the *Electronic Discovery Reference Model* found at EDRM.net.

[3] For an example of this industry standard two-step practice see *Gabriel Techs., Corp. v. Qualcomm, Inc.*, No. 08CV1992 AJB (MDD), 2013 WL 410103 at *10 (S.D. Cal. Feb. 1, 2013) (cost award allowed to prevailing party of $2,829,349.10 for first-pass review by SMEs of one million documents and another $391,928.91 for a second review by contract lawyers.)

[4] It is included in the *EDBP* as step seven and described as *Computer Assisted Review* (CAR) found at: http://www.edbp.com/search-review/computer-assisted-review/.

[5] It is included in the *EDBP* as step eight and described as *Protections* found at: http://www.edbp.com/search-review/.

[6] *Mt. Hawley Ins. Co. v. Felman Production, Inc.*, 2010 WL 1990555 (S.D. W. Va. May 18, 2010) (privilege documents accidentally produced resulting in waiver of privilege in

Engineering, Inc., a large construction case involving millions of documents reviewed for possible production, both sides inadvertently produced thousands of privileged documents to each other.[7] They did so despite expenditures of tens of millions of dollars for traditional attorney review of each document before production.

My analysis and experiments with cost controls since 2006 have focused on both the cost of the initial relevancy review, and the cost of the final protection reviews. Great progress has been made in the last several years in lowering the cost of first-pass relevancy reviews, especially when using artificial intelligence ("AI") enhanced software. Improvements have also been made in the second review costs, but not nearly as dramatic. Although some advocate for the elimination of second pass review to save expenses, and reliance on confidentiality and clawback agreements only for protection of confidential client documents, most litigants today are unwilling to take the risks involved.

At the present time, in most cases, virtually no corporate clients are willing to dispense with final manual review of documents selected for production and rely *solely* on automated software for protections. The likelihood of error is simply still too high for this to be an acceptable risk in most cases for most clients. The damage caused by disclosure of some privileged communications cannot be fully repaired by *clawback* agreements.[8]

---

spite of sophisticated counsel employing elaborate safeguards). *Also see Diabetes Centers of America, Inc. v. Healthpia America, Inc.,* 2008 U.S. Dist. LEXIS 8362, 2008 WL 336382 (S.D. Tex. Feb. 5, 2008) (production errors made by both sides); *Danis v. USN Communications, Inc.*, 2000 WL 1694325 (N.D. Ill. 2000) ($10,000 fine imposed against CEO personally when the young general counsel he hired to supervise e-discovery was grossly negligent);

[7] *Tampa Bay Water v. HDR Engineering, Inc.* Case No. 8:08-CV-2446-T-27TBM. (M.D. Fl. November 2, 2012) (also found at 2012 U.S. Dist. LEXIS 157631 and 2012 WL 5387830). The plaintiff alone inadvertently produced 23,000 privileged documents. The prevailing defendant in this case was awarded over twenty million dollars in fees and costs. Of this sum $3,100,000 was awarded as a cost for e-discovery vendor processing and hosting of 2.7 million documents for review. Another $4,590,000 ($1.70 per file) is estimated to have been spent by one defendant in *attorney fees* to review the documents. *See* Losey, R., *$3.1 Million e-Discovery Vendor Fee Was Reasonable in a $30 Million Case* (e-Discovery Team, Aug. 4, 2013) found at http://e-discoveryteam.com/2013/08/04/3-1-million-e-discovery-vendor-fee-was-reasonable-in-a-30-million-case/#comment-60139.

[8] *See:* FN 4. *Also see: Brookfield Asset Management*, Inc. v. AIG Financial Products Corp., 2013 WL 142503 (S.D.N.Y. Jan. 7, 2013); Losey, R., *Another Clawback*

As I explained in my series *Secrets of Search,* Parts One, Two and Three, the latest AI enhanced software is far better than keyword search, but not yet good enough to allow for a fully automated approach to protection review.[9] Based on my informal surveys and discussions with attorneys around the country, I have found that confidentiality reviews are only omitted in certain non-litigation circumstances, such as when making corporate merger related productions to the government, or in cases where the data under review is very unlikely to contain confidential information, such as old data of an acquired company. Also, I have seen it done in bankruptcy cases, or by litigants in any type of case who otherwise simply could not afford to respond properly to discovery requests.

Although AI enhanced review is not good enough, *yet*, to rely on exclusively for protection review, it is certainly up to the task of first pass relevancy review.[10] Attorneys should also use AI enhanced search to *speed up* the second pass manual review for protections.

Of course, it is the litigant's data and their confidential information, so if a litigant (or third party responding to a subpoena), really does not care about disclosure of confidential information in any particular data set, and is not required by law to protect the confidential information, and if they want to rely on claw back orders and confidentiality agreements alone, then that is their right. They may instruct their attorneys to skip this step for cost saving purposes. If you are an attorney receiving such an instruction, I recommend that you confirm that instruction in writing. You should also provide a full, detailed, written disclosure of the risks involved. Make sure the instruction is based on the client's full understanding that once a bell has been rung, it cannot be un-rung, and that, for instance, waiver of once privileged attorney-client communications may open the door to waiver of others.

Enforcement Order Shows the Importance of the Selection of Quality Vendors found at http://e-discoveryteam.com/2013/03/12/another-clawback-enforcement-order-shows-the-importance-of-the-selection-of-quality-vendors/

[9] The three part essay on Search is available online in one document found at: http://ralphlosey.files.wordpress.com/2013/03/secrets_of_search_2012_consolidated.pdf

[10] For a description of what I mean by *AI enhanced search*, which is also commonly called *Predictive Coding, Technology Assisted Review (TAR), or Computer Assisted Review (CAR)* see my page describing CAR found at http://e-discoveryteam.com/car/. *Also see* the descriptions for step seven in the EDBP at http://www.edbp.com/search-review/.

The Idea of *Bottom Line Driven Review*

After two years of analysis of the review cost problem I came up with an idea in 2008 that looked promising. It is simple in concept, which is one of its strengths (although its implementation can sometimes be complex). I have since tested and refined this method in two law firms with multiple types of cases and investigations. I have also spoken about this approach with many other attorneys, judges, and law professors, and taught this method at multiple CLE events around the country.[11] I call it **Bottom Line Driven Proportional Review and Production**. A more technical description for it, the one I used in a legal methods patent application, is: *System and Method for Establishing, Managing, and Controlling the Time, Cost, and Quality of Information Retrieval and Production in Electronic Discovery.* But I usually just call it *Bottom Line Driven Review*.

The Bottom Line of Productions

The bottom line in e-discovery production is what it costs. Believe me fellow lawyers, clients care about that …. a lot! In *Bottom Line Driven Proportional Review* everything starts with the bottom line. What is the production going to cost? Despite what some lawyers and vendors may tell you, it is not an impossible question to answer. It takes an experienced lawyer's skill to answer, but after a while, you can get quite good at such estimation. It is basically a matter of man-hours estimation. With my method it becomes a reliable art that you can count on. It may never be exact, but the ranges can usually be predicted, subject of course to the target changing after

---

[11] I am going to explain the idea here with enough detail for attorneys experienced in e-discovery to be able to try it out. I urge you to do so. I have been working and training the attorneys in my current law firm, Jackson Lewis, in this for years. We know that it works in all types of employment law cases, federal and state, large and small, especially when coupled with the *doctrine of proportionality* (also explained in this essay). In my prior law firm, Akerman Senterfitt, the method was tested and refined in multiple types of commercial litigation. Time and conflict checks permitting, I can consult and assist attorneys outside of my firm on *Bottom Line Driven Review*, especially when it involves the use of *artificial intelligence* enhanced software (often called *predictive coding*) to search for evidence in *big data*. That is my current area of special interest and *Bottom Line Driven Review* is particularly effective in large cases.

the estimate is given. If the complaint is amended, or different evidence becomes relevant, then a *change order* may be required for the new specifications.

Price estimation is second nature to me, and an obvious thing to do before you begin work on any big project. I think that is primarily because I worked as a construction estimator out of college to save up money for law school back in the seventies. Estimating legal review costs is basically the same thing, projecting materials and labor costs. In construction you come up with prices per square foot. In e-discovery you estimate prices per file, as I will explain in detail in this essay.

My new strategy and methodology is based on the bottom line. It is based on projected review costs, defensible culling, and best practices of AI enhanced review. Under this method the producing party determines the number of documents to be subjected to costly reviews by calculating backwards from the bottom line of what they are willing, or required, to pay for the production.

Setting a Budget Proportional to the Case

The process begins by the producing party calculating the maximum amount of money appropriate to spend on the production. A budget. This requires not only an understanding of the production requests, but also a careful evaluation of the value and merits of the case. This is where the all important proportionality element comes in.

The amount selected for the budget should be proportional to the monies and issues in the case. Any more than that is unduly burdensome and prohibited under Rule 26(b)(2)(C), *Federal Rules of Civil Procedure* and other rules that underlie what is now generally known as the *Proportionality Principle*.[12]

---

[12] *See* Rule 1, Rule 26(b)(2)(C), Rule 26(b)(2)(B), and Rule 26(g), *Federal Rules of Civil Procedure*; *Commentary on Proportionality in Electronic Discovery*, 11 SEDONA CONF. J. 289 (2010); Carroll, *Proportionality in Discovery: A Cautionary Tale,* 32 Campbell L. Rev. 455, 460 (2010); Losey, R. *Good, Better, Best: a Tale of Three Proportionality Cases – Part Two* found at http://e-discoveryteam.com/2012/04/15/good-better-best-a-tale-of-three-proportionality-cases-part-two/. *Also see*: *Tamburo v. Dworkin*, 2010 WL 4867346 (N.D. Ill. Nov. 17, 2010); *Rimkus Consulting Group v. Cammarata*, 688 F.Supp. 2d 598, 613 (S.D. Tx. 2010); *Apple v. Samsung*, 2013 WL 442365412 (N.D. Cal. August 14, 2013); *Wood v. Capital One Services, LLC,* No. 5:09-CV-1445 (NPM/DEP), 2011 WL 2154279, at *1-3, *7 (N.D.N.Y, 2011); *Daugherty v. Murphy*, No. 1:06-cv-0878-SEB-DML, 2010 WL 4877720, at *5 (S.D. Ind., 2010); *Willnerd v. Sybase*, 2010 U.S. Dist. LEXIS 121658 (SD Id., 2010); *I-Med Pharma Inc. v. Biomatrix,* 2011 WL 6140658 (D.N.J. Dec. 9, 2011); *U.S. ex rel McBride v. Halliburton Co.*,, 272 F.R.D. 235 (D.D.C.

The budget becomes the bottom line that drives the review and keeps the costs proportional. The producing party seeks to keep the total costs within that budget. The budget should either be by agreement of the parties, or at least without objection, or by court order.

The failure to estimate and project future costs, and to plan and limit reviews so that they stay within budget, accounts for much of today's out of control e-discovery costs. Once you spend the money, it is very hard to have costs shifted to the requesting party. But if you raise objections and argue proportionality *before the spend*, then you will have a fair chance to constrain your expenses within a reasonable budget.

Under the *Bottom Line Driven Proportiona*l approach, after analysis of the case merits, and determination of the maximum expense for production proportional to a case, the responding party makes a good faith estimate of the likely maximum number of documents that can be reviewed within that budget. The document count represents the number of documents that can be reviewed for final decisions of relevance, confidentiality, privilege and other issues, and still remain within budget. The review costs you estimate must be based on best practices and be accurate (no puffing).

Following best practices the producing party then uses advanced search techniques and quality controls to find the documents in a first pass review that are most likely to be relevant and stay within the number of documents allowed by the budget. Since predictive coding type AI enhanced software ranks all documents according to probable probative value, it the perfect tool to facilitate *bottom line driven review*.[13] The ranking feature makes it far easier to focus on the most important documents and stay within budget.

Good predictive coding software today evaluates the strength of the relevance and irrelevance of every document in the data set analyzed. That is one reason I

2011); *DCG Sys., Inc. v. Checkpoint Techs, LLC*, 2011 WL 5244356 (N.D. Cal. Nov. 2, 2011).

[13] Losey, R. *Relevancy Ranking is the Key Feature of Predictive Coding Software* found at http://e-discoveryteam.com/2013/08/25/relevancy-ranking-is-the-key-feature-of-predictive-coding-software/
Unfortunately, the use of AI software comes with its own transactional costs, which means it cannot be economically used in cases that are too "small." Typically this means cases involving less than a $25,000 budget for document review. For these smaller cases the same bottom line approach should still be used, and it can still work fairly well, even if you do not have the benefit of expensive AI software and its ranking properties.

was especially pleased when AI type software with reliable ranking abilities first came into the market in 2011 and have since started specializing in the best methods for use of this software. Bottom line driven review was, and still can be, done without predictive coding ranking (I did so for years), but it is harder to be accurate without computerized ranking.

Using best methods and AI search with relevancy ranking allows you to get the most *bang for your buck*, the core truth. This in turn helps persuade the requesting party (or the court, should agreement not be reached), to go along with your proposed budget constraints.

Unfortunately, the use of AI software comes with its own transactional costs, which means it cannot be economically used in cases that are too "small." Typically this means cases involving less than a $25,000 budget for document review. For these smaller cases the same bottom line approach should still be used, and it can still work fairly well, even if you do not have the benefit of expensive AI software and its ranking properties.

"Small" Document Review Example

A few examples should help clarify how this works. If you set a proportional document review cost for a case of $25,000.00, and estimate based on experience, sampling and other hard facts that it will cost you about $1.00 per file for both the automated first-pass and subsequent manual review of documents, then you can review no more than 25,000 documents and stay within budget. It is basically that simple. No higher math is required.

The most difficult part is the legal analysis to determine a budget proportional to the real merits of the case. But that is nothing new. What is the *golden mean* in litigation expense? How to balance *just*, with *speedy* and *inexpensive*?[14] The ideal proportionality question of perfect balance has preoccupied lawyers for decades. It has also preoccupied scientists, mathematicians, and artists for centuries. Unlike lawyers, they claim to have found an answer, which they call the *golden mean* or *golden ratio*:

$$a \qquad b$$

$$a+b$$

$a+b$ is to a as a is to b

---
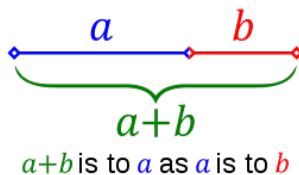
[14] Rule One of the Federal Rules of Civil Procedure requires all other rules to be interpreted to accomplish the *just, speedy* and *inexpensive* resolution of all cases.

In law this is the perennial *Goldilocks* question. How much is too much? Too little? Just right? How much is an appropriate spend for document production? The issue is old. I have personally been dealing with this problem for over thirty-three years. What is new is applying that legal analysis to a modern-day high-volume electronic document search and review plan. Unfortunately, unlike art and math, there is no accepted *golden ratio* in the law, so it has to be recalculated and reargued for each case.[15]

Estimation for bottom line driven review is essentially a method for marshaling evidence to support an undue burden argument under Rule 26(b)(2)(C). It is basically the same thing we have been doing to support motions for protective orders in the paper production world for over sixty years. The only difference is that now the facts are technological, the numbers and varieties of documents are enormous, sometimes astronomical, and the methods of review are very complex and not yet standardized.

Estimate of Projected Costs

The calculation of projected cost per file to review can be quite complicated, and is frequently misunderstood, or is not based on best practices. Still, in essence this cost projection is fairly simple. You basically project how long it will take to do the review and the total cost of the time. To be more accurate you can also include the *materials* costs, for example, software usage fees, processing costs, and other non-legal vendor charges.[16]

---

[15] If the golden ratio were accepted in law as an ideal proportionality, the number is 1.61803399, aka *Phi*. That would mean 38% is the perfect proportion. I have argued that when applied to litigation that means the total cost of litigation should never exceed 38% of the amount at issue. In turn, the total cost of discovery should not exceed 38% of the total litigation cost, and the cost of document production should not exceed 38% of the total costs of discovery (as opposed to our current 73% reality). (It's like Russian nesting dolls that get proportionally smaller.) Thus for a $1 million case you should not spend more than $54,872 for document productions (1,000,000 – 380,000 – 144,400 – 54,872). *See* Losey, R., *Beware of the ESI-discovery-tail wagging the poor old merits-of-the-dispute dog*. I have not yet made this math and art golden ratio argument in court, but, who knows, it just might work with the right judge. To me the proportions seem reasonable.

[16] For a full picture of total e-discovery spend you should include other e-discovery costs unrelated to review, such as vendor and attorney costs related to preservation and

Thus, for example, think of our relatively small project to review 25,000 documents.[17] Your first step in the review is to identify the relevant documents in these 25,000, or put another way, to weed-out the irrelevant documents. In most types of lawsuits, less than five percent of a custodian's emails will be relevant. Sometimes it is less than one percent, or a tenth of a percent. The primary task is to quickly and efficiently identify the rare relevant needles from the large haystacks of irrelevant documents. Once the probable irrelevant documents are removed and the probable relevant documents are identified, the next *Protections* review stage can begin. As mentioned, documents coded relevant, and only those documents, are then re-reviewed for privilege and confidentiality, and redacted, labeled and logged as necessary. They are often also issue tagged at this stage for the later use and convenience of trial lawyers, or sometimes the issue tagging may be included in the first pass review. Mistakes in first pass relevancy review are also corrected with quality controls built into the reviewer classifications.

First pass relevancy reviews were originally done by having a lawyer actually look at, meaning skim or read, each of the 25,000 documents. Using low paid contract lawyers, this kind of first-pass relevancy review typically goes at a rate of from between 50 to 100 files per hour. But by using AI enhanced review software, a skilled search expert, who must also be a *subject matter expert* (SME) for predictive coding to work well, can attain speeds in excess of 1,000 files per hour (or even much faster than that when larger volumes of documents are involved) for first pass review.[18]

Very Small Productions

---

collection. Recall the Rand survey showed found these non-review related expenses to average about 27% of the total e-discovery costs.

[17] Note it probably started as 100,000 documents collected for preservation and possible review, but you bulk-culled it down to 40,000 documents for review (*Culling* is step six in the EDBP). Bulk culling is performed by making such legal decisions as custodian ranking (for instance, a decision to review only the top five most important custodians, even though the email of ten custodians was collected), date ranges, and file types (for instance, exclude all music and photo files types). *See* http://www.edbp.com/search-review/bulk-culling/.

[18] For more on how the predictive coding review process works, including the importance of qualified SMEs to train the machine, a process called active machine learning, see my introduction to the subject at http://e-discoveryteam.com/car/ and http://www.edbp.com/search-review/computer-assisted-review/ and other articles cited therein.

When you are working with a volume of approximately 25,000 or more documents, as in this example, then the AI enhanced SME type of review is usually a financially viable approach. But with fewer documents than that, or a smaller budget, it may not be. The exact cut off point depends on the costs you incur with your vendor for use of the software, and the costs (and sometimes availability) to use a qualified SME to do the AI enhanced reviews. It also depends on the value of the case.

For very small document productions, but ones that still involve more than a few thousand documents, the linear review of each and every document is still not the solution. Document by document review of thousands of documents by multiple reviewers is a notoriously inefficient, ineffective, and costly approach. Instead, we fall back on other methods for first pass culling that were *state of the art* just a few years ago before AI enhanced software became available. These easier and less expensive methods employ what I call a *multimodal tested keyword* approach. The vast majority of the employment law cases I deal with now involve very small productions where we use this alternative non-AI approach.

The methods rely primarily on tested keyword searches. Tested keyword search is described in part by Judge Peck's well-known *wake up call* opinion in *Gross Construction*, which provides basic advice to lawyers on the proper way to use keyword search.[19] In essence, you do not simply *guess* what key words might be effective, but instead you use witness interviews and other factual research and analysis to derive various possible terms. You then test proposed keywords using both judgmental and random sampling. You also use keywords with Boolean combinations and parametric field limitations (applied to various metadata fields).

Other types of search, including similarity searches, concept searches, and even some linear review of key time periods and custodians, are used to supplement the tested parametric Boolean keyword searches. This is why I call it a *multimodal* method. This type of search was *state of the art* just a few years ago. It still has the advantage of being easier to use, requiring less sophisticated/expensive software, and far less SME involvement. But, compared

---

[19] *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009). *Also see*: Losey, R. *Child's Game of "Go Fish" is a Poor Model for e-Discovery Search* found at http://e-discoveryteam.com/2009/10/04/childs-game-of-go-fish-is-a-poor-model-for-e-discovery-search/.

to AI enhanced review it is far less reliable, and when large volumes of documents are involved, far more expensive.

Relative Speeds of Review

Assuming your document review project is large enough, and you can use AI enhanced software and good SMEs, you can determine file relevancy at speeds hundred of times faster than linear review by contract lawyers, and can do so far more accurately. That is why an SME with good software can charge ten to twenty times as much as a contract lawyer and still do the first pass review at a fraction of the cost. There is a recent example of this cost-savings approach in *Gabriel Techs. Corp. v. Qualcomm, Inc.,* a 2013 patent case in District Court in California that considered the reasonability of a fee award to a prevailing party.[20] A fee of $2,829,349.10 was awarded for the work of a team of SMEs to review one million documents for possible relevance.

Although I cannot discuss the actual cases I have handled, in 2012 I did a public experiment using what I call a multimodal AI enhanced type of review of 699,082 ENRON documents. I was looking for evidence in this public ENRON dataset concerning involuntary employee terminations, a fairly simple relevancy determination. I completed the first pass review at an average speed of 13,444 files per hour.[21] Speeds such as this are common in the type of employment law issues I have been dealing with in the past couple of years, but similar speeds are attainable in other types of cases too. For instance, I recently completed another more complex first-pass review of over a million and a half documents at an average speed of 35,831 files per hour. I did this review myself, since I happened to be the only SME available for this project involving non-employment law legal issues. There were significant time pressures for this review and I completed the first-pass review of over a million and a half documents in only two weeks.

Back to the Small Case Example

Back to the small case example of only 25,000 documents, let us assume a modest AI enhanced first-pass review speed of 2,000 files per hour. This means

---

[20] *Gabriel Techs. Corp. v. Qualcomm, Inc.*, No. 08CV1992 AJB (MDD), 2013 WL 410103 (S.D. Cal. Feb. 1, 2013) (keyword bulk culled 12,000,000 documents down to 1,000,000 (not a best practice) and then used predictive coding type rule based algorithms (not AI) to determine relevancy of the remaining 1,000,000 relevant).

[21] *See Predictive Coding Narrative: Searching for Relevance in the Ashes of Enron* found at http://ralphlosey.files.wordpress.com/2013/04/predictive-coding-narrative_corrrected_3-21-13.pdf

an SME could complete the review in 12.5 hours. (Typically SME's using AI enhanced software review far more than 25,000 documents and so are able to attain faster speeds. The math and costs savings still work if the cost of using the predictive coding software and other transaction costs are not too high.) It would probably take the SME about 2.5 hours to master the particular factual issues in the case, so let us assume a total time of 15 hours, and a review rate for this SME of $500.00 per hour (in a small case like this SMEs at relatively low rates like this are common, whereas in larger cases the SME rates can be much higher, but the speed of review and savings realized can also be much larger.) That means an expense for first pass review (excluding software charges) of $7,500.00, which is still half the cost of traditional manual review.

Under a traditional contract lawyer review, where we assume a very fast speed (for them) of 75 files per hour (producing a total of 333.33 hours of review time), and a very low *unmotivated* lawyer rate of $50.00 per hour, you have a projected fee of $16,666.50. Even though the $500 per hour rate of the SME is 10-times higher than the contract lawyers, since the SME is 26.67 times faster, the SME AI enhanced review still costs half as much. That is because it would take the contract lawyers 333.33 hours to complete the project, and, this is important, they would necessarily do so with a far **lower** *accuracy* rate. The contract lawyers working without the assistance of AI augmentation are likely to find *far fewer* relevant documents than the automated SME approach.[22] This makes clear the power and importance of SMEs driving predictive coding review, and why, along

---

[22] To back up the likely far better *accuracy* claim for an SME driving a CAR, please see the following scientific studies: Grossman & Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review,* Rich. J.L. & Tech., Spring 2011; Grossman & Cormack, *Inconsistent Responsiveness Determination;* Roitblat, Kershaw, and Oot, *Document categorization in legal electronic discovery: computer classification vs. manual review*, Journal of the American Society for Information Science and Technology, 61(1):70–80, 2010; Voorhees,*Variations in relevance judgments and the measurement of retrieval effectiveness*, 36:5 Information Processing & Management 697, 701 (2000); Losey, R., *A Modest Contribution to the Science of Search: Report and Analysis of Inconsistent Classifications in Two Predictive Coding Reviews of 699,082 Enron Documents*, found in two parts at http://e-discoveryteam.com/2013/06/11/a-modest-contribution-to-the-science-of-search-report-and-analysis-of-inconsistent-classifications-in-two-predictive-coding-reviews-of-699082-enron-documents/, and http://e-discoveryteam.com/2013/06/17/comparative-efficacy-of-two-predictive-coding-reviews-of-699082-enron-documents/.

with their current scarcity, attorneys with both subject matter expertise and advanced search skills are now in such demand.[23]

In our hypothetical after the first pass review of 25,000 documents the slower protection review begins. Now the highly skilled SMEs are no longer required. The lower-paid contract lawyers, or associates and paralegals, can do the review of the documents the SMEs have determined to be relevant. (There is still, of course, expert attorney supervision required of the contract reviewers, associates and paralegals.) Let us assume that the first pass review found that 2,500 of the 25,000 documents were relevant. (The ten percent relevance ratio, aka "prevalence rate," is a relatively high rate.) This means that only 2,500 documents are subject to second pass review for confidentiality. Let us assume this work goes at an average rate of 20 files per hour. (Remember this review includes incidental related activities such as preparing a privilege log.) The 2,500 documents could be reviewed for protections in 125 hours at a cost of $6,250.00. So our *base minimum* review cost for both passes is $13,750.00.

I say *base minimum* because you have additional expenses beyond just contract reviewer time, or junior associate and paralegal time, including the expense of partner, senior associate, and project manager time to supervise other attorneys and paralegals, perform quality control reviews, etc., plus software and hosting costs. Let us assume that is another $7,000.00 cost here, for a total expense of $20,750.00.

Congratulations! You came in under your $25,000.00 budget. You completed your review of 25,000 documents at a cost of $0.83 per document instead of the $1.00 per document you had estimated. For smaller projects of less than 25,000 documents, where AI enhanced review would probably not be cost effective, the per document estimate would be higher. Still, a multimodal tested keyword approach can achieve substantial savings and better quality than a brute force linear "review everything" approach.

That $0.83 per document in this small case hypothetical is pretty good. But in larger projects, where you are talking about millions of documents to review, and more realistic prevalence rates, where frequently less than five percent of the

---

[23] *See* David Cowen, Job Market Heating Up for e-Discovery Technologists, Managers, and Attorneys; Losey, R., Reinventing the Wheel: My Discovery of Scientific Support for "Hybrid Multimodal" Search found at http://e-discoveryteam.com/2013/04/21/reinventing-the-wheel-my-discovery-of-scientific-support-for-hybrid-multimodal-search/.

total documents are relevant and thus need to be second reviewed, the savings can be even greater. With larger document volumes the review speeds are much higher, and per document review costs are lower, sometimes, much lower.

Estimation Caveats

You could estimate all of these costs in advance by having a bank of experience to draw upon, knowing the likely costs per file range. But remember, even in the world of repeat litigation, like employment law claims, all projects are different. All document sets are different. You have to, as I like to say, *get your hands dirty in the digital mud*. You have to know your ESI collection before you make an estimate.

Even in just one *type* of electronic document, the one most common in e-discovery today, email and attachments, the variances in email collections can be tremendous. With those variables come different review speeds and confidentiality concerns. The review speeds, and thus the review costs, depend on the density of the documents themselves and difficulty of the document classifications. That is where the art of estimation comes in and depends on your understanding of the project. You have to understand the blueprints and specs of a project before you can make a valid estimation.

This is especially true of the SMEs' first-pass work. You need to do some sampling to see what review rates apply. How long will it take these particular SMEs to do the tasks assigned to them in this case with this data. Sampling is the only reliable way to do that, especially when it comes to the all important prevalence calculations. As a general rule of thumb I have found that the higher the prevalence, the lower the review speed and thus higher the cost. Random sampling to determine prevalence is not only a routine part of my normal eight-step AI-enhanced review process to ensure quality control,[24] but also an important step to give reliable cost estimates.

---

[24] A random sample is step-three in my recommended eight-step methodology for AI enhanced review, aka *predictive coding*. *See eg* my blog at http://e-discoveryteam.com/car/ and the EDBP at http://www.edbp.com/search-review/predictive-coding/. For more on random sampling *see* Losey, R., *Comparative Efficacy of Two Predictive Coding Reviews of 699,082 Enron Documents* (Part Two) found at http://e-discoveryteam.com/2013/06/17/comparative-efficacy-of-two-predictive-coding-reviews-of-699082-enron-documents/; and Losey, R. Random Sample Calculations And My Prediction That 300,000 Lawyers Will Be Using Random Sampling By 2022 found at http://e-discoveryteam.com/2012/05/06/random-sample-calculations-and-my-prediction-that-300000-lawyers-will-be-using-random-sampling-by-2022/.

A Big Data Example

Let us change the scenario somewhat for a final quick hypothetical. Assume **10,000,000 documents** *after bulk culling* for your SMEs to review for relevance. Assume sampling by an SME showed a prevalence of 10% (again somewhat high), and a first-pass review rate of 10,000 files per hour (somewhat slow for Big Data reviews). This means you can assume that only around 1,000,000 documents will need final second pass protection review.

More sampling shows the contract reviewers will, using computer-assisted techniques (smart routing, etc.), be able to review them at 100 files per hour. The improvement in protections review from the small case example, where a speed of only 20 files per hour was assumed, is possible when you have larger document collections sizes like this (one million documents), high quality review software, and experienced review teams.

With this information from sampling you can now estimate a total first pass review cost of $1,000,000.00 ($1,000.00 x 1,000 hours), which compares pretty well to the $2,829,349.10 charge in *Gabriel.* Next you can estimate a total final pass Protection review cost of $500,000.00 ($50 x 10,000 hours).

Lets us also assume you know from experience that your other supervision fees and software costs are likely to total another $150,000.00. Your estimate for the project would be $1,650,000.00. That is 16.5 cents a document ($0.165).

Too high you say? Not proportionate to the value of the case? Not proportionate to the expected probative value in this case from these 10,000,000 documents, which is something your sampling can tell you and provide evidence to support? Then use the document ranking feature of AI enhanced document review to further reduce the second pass protection costs.[25]

If the SME's identification of 1,000,000 likely relevant documents was based on a 50% plus probability ranking using predictive coding, then try a higher-ranking cut off. Again, this is not too difficult to do with sampling and good software. Maybe a 75% plus ranking cut off brings the count down from 1,000,000 to 250,000. Or

---

[25] Losey, R. *Relevancy Ranking is the Key Feature of Predictive Coding Software* found at http://e-discoveryteam.com/2013/08/25/relevancy-ranking-is-the-key-feature-of-predictive-coding-software/. Relevancy ranking only works well now with the best software on the market and requires proper, accurate, machine training. Even then it is not an exact measure of relevancy weight, and testing and quality controls should always be applied to know when and to what degree the ranking strata are reliable.

maybe you just arbitrarily decide to use the top ranked 250,000 documents because that is all you can afford, or think is proportionate for this data and this case. Either way, you are now passing the strongest document sets along for second pass review. You will probably be producing the most relevant documents, the ones likely to have the highest probative value to the case.

Using the higher cut off, assume that the cost for second pass Protection review would then be 25% of what it was, reduced from $500,000.00 for review of 1,000,000 documents, to $125,000.00 to review 250,000 documents. Your *other fees and costs* also drop in your experience by 50%, from $150,000.00 to $75,000.00. Your total estimate is now $1,200,000.00, instead of $1,650,000.00. You have gone down to twelve cents a document. Assume this $1,200,000.00 number is now within your legal judgment to be proportional to this document request in this very large lawsuit. It is now within your budget.

If this was still not proportional in your judgment, or the client's opinion, there are many other things you could try, but I would focus on the bulk culling before the SME first pass relevancy review. For instance, change the custodian count, or date range, but do not keyword filter like they did in *Gabriel*. Bring the initial 10,000,000 documents down to 5,000,000 documents, then do the math (500,000+250,000+37,500). You are now talking about around $800,000.00, back to sixteen cents per document. Is that within the budget? Is that an amount that a court is likely to force you to spend anyway?[26]

---

[26] The costs of review have come way, way down in the past few years for those who are using advanced methods. For some context on the fourteen cents a document number used in this example, back in 2007 the Department of Justice spent $9.09 per *document* for review in the *Fannie Mae* case, even though it used contract lawyers for the review work. *In re Fannie Mae Securities Litig*., 552 F.3d 814, 817 (D.C. Cir. 2009) ($6,000,000/660,000 emails). There were no comments by the court that this price was excessive when the government later came back and sought cost shifting. At about the same time Verizon paid $6.09 per document for a massive second review project that enjoyed large economies of scale and, again, utilized contract review lawyers. Roitblat, Kershaw, and Oot, *Document categorization in legal electronic discovery: computer classification vs. manual review*. Journal of the American Society for Information Science and Technology, 61(1):70–80, 2010 ($14,000,000 to review 2.3 million documents in four months). In 2011 I was still seeing an average cost of $5.00 per file for tested keyword multimodal reviews before advanced predictive coding engines started to become available.

The *More Bang for the Buck* Bottom Line Ranked Approach is Good for Both the Requesting Party and the Producing Party

When a producing party uses advanced computer review software with predictive coding ranking, they can limit their review and production to the documents with the highest relevancy ranking. This gives everyone the most truth for the dollar. For that reason this is the best approach, not only for the producing party, who can thereby budget and avoid disproportionate burdens, but also for the requesting party.[27] They can also limit their expense to review the many thousands of documents (sometimes millions of documents) produced to them. They can also focus on the documents with the highest value. They can be reasonably assured that AI enhanced processes have probably found the documents that really matter to the resolution of the case. The plaintiffs understood in *Kleen Products*, a large class action pending in District Court in Chicago, which is one reason they tried to force the defendants to use predictive coding in their productions.[28]

In spite of the *Kleen Products* precedent, not all plaintiff's counsel are as sophisticated as counsel in that large class action. A producing party, especially in a *David and Goliath* type case, will often have to *sell* the benefits of these new methods to the smaller requesting party. The requesting party will be more likely to cooperate if they understand the value to them of these methods. This will often require the producing party to provide some reasonable degree of transparency into the proposed review processes. For instance, if you have well known experts retained to direct the review, tell them. Share his or her qualifications and experience. Engage in frank and full relevancy dialogues with the requesting party.[29] Make sure you are training the machine to find the documents that they really want. Clarify the target. If need be, share some examples early on of the relevant documents you will use in the training. Invite them to provide you with documents they consider relevant to use in the training.

---

[27] See: Losey, R., *Why a Receiving Party Would Want to Use Predictive Coding?* found at http://e-discoveryteam.com/2013/08/12/why-a-receiving-party-would-want-to-use-predictive-coding/.

[28] *Kleen Products, LLC, et al. v. Packaging Corp. of Amer., et al.,* Case: 1:10-cv-05711, Document #412 (ND, Ill., Sept. 28, 2012).

[29] Relevancy dialogues, which I also call *ESI Discovery Communications*, is step-one in my standard eight step methodology for predictive coding. See the EDBP description found at: http://www.edbp.com/search-review/predictive-coding/.

Try to make it a collaborative approach.[30] If they refuse, document your efforts and take any disputes to the judge as early as possible.

A collaborative approach is an effective way to convince a requesting party that your estimate is reasonable and that you are not still stuck in the old paradigm of *hide-the-ball* discovery games. I cannot overstate how important it is to develop *trust* between opposing counsel on discovery. Often the only way to do that is through some level of transparency. You do not have to disclose all of your search secrets, but you may have to keep the requesting party informed and involved in the process. That is what cooperation looks like.

All Review Projects Are Different

In order to be able to make a valid estimate for *bottom line driven proportional review* you must closely study the case and review the project goals. It is not enough to have handy per file cost estimates. This move to actual examination of the electronically stored information at issue, and study of the specific review tasks, is equivalent to the move in construction estimation from rough estimates based on average per square foot prices, to a site visit with inspection and measurements of all relevant conditions. No builder would bid on a project without first doing the detailed real world estimation work. We lawyers must do the same for this method to succeed.

Even in the same organization, and just dealing with email, the variances between document custodians can be tremendous. Some, for instance, may have large amounts of privileged communications. This kind of email takes the most time to review, and, if relevant, to log. High percentages of confidential documents, especially partially confidential, can also significantly drive up the costs of the second pass review. All of the many unique characteristics of electronic document collections can affect the speed of review and total costs of review. That is why you have to look at your data and test sample the emails in your collection to make accurate predictions. Estimation in the blind is never adequate. It would be like bidding on a construction project without first reading the plans and specifications.

---

[30] *Kleen Products, LLC, et al. v. Packaging Corp. of Amer., et al., 2012 WL 4498465* (ND, Ill., Sept. 28, 2012). *Also see:* Losey, R., Attorneys Admonished by Judge Nolan Not to "Confuse Advocacy with Adversarial Conduct" and Instructed on the Proportionality Doctrine.

Even when you have dealt with a particular client's email collection before, a repeat customer so to speak, the estimates can still vary widely depending on the type of lawsuit, the issues, and the amount of money in controversy or general importance of the case.

Although this may seem counter-intuitive, the truth is, complex, big-ticket cases are the easiest to do e-discovery, especially if your goal is to do so in a proportional manner. If there is a billion dollars at issue, a reasonable budget for document review is large. On the other hand, proportional e-discovery in small cases is a real challenge, no matter how simple they supposedly are. Many cases that are small in monetary value are still very complex. And complex or not, all cases today have a lot of electronically stored information.

The medium size to small cases are where my *bottom line driven proportional review* has the highest application for cost control and the greatest promise to bring e-discovery to the masses. Give it a try, with our without AI enhanced predictive coding. For large cases, you should <u>always</u> employ AI enhanced methods. The ranking feature makes it easier to implement a proportional approach. *Bottom line driven proportional review* in big cases can save the greatest amount of money and provide much needed certainty and risk control.

Conclusion

The future of litigation involves new methods of technology-assisted discovery where Man and Machine work together to find the core truth. This day will come; in fact it is already here. As the science fiction writer William Gibson said: *The future is already here – it's just not evenly distributed yet*. The key facts you need to try a case and do justice can be found in any size case, big and small, at an affordable price. But you have to embrace change and adopt new legal and technical methodologies. The *Bottom Line Driven Proportional Review* method is part of that answer, so too is advanced review software at affordable prices. It is already working for some today, it can work for you too.

For questions or comments, please feel free to contact the author at: Ralph.Losey@JacksonLewis.com.