

# More on the Ethics of E-Discovery: Predictive Coding and Other Forms of Computer-Assisted Review

**John M. Barkett**  
**Shook, Hardy & Bacon L.L.P.**  
**Miami, Florida**

<b>Introduction - An Increasingly Realistic Hypothetical .....</b>	<b>3</b>
<b>Case Law Calling for Cooperation .....</b>	<b>5</b>
<b>A Duty to Cooperate in Rule 1? .....</b>	<b>6</b>
<b>Case Law on Cooperation and Search Technology .....</b>	<b>7</b>
<b>The Perils of Search Ignorance .....</b>	<b>10</b>
<b>The Legal Track at TREC .....</b>	<b>11</b>
<b>Da Silva Moore .....</b>	<b>20</b>
Custodians .....	21
Sources of ESI .....	21
Predictive Coding Protocol .....	22
Lessons for the Future .....	23
The District Court’s Adoption of Judge Peck’s February 24, 2012 Order .....	25
<b>Kleen Products .....</b>	<b>27</b>
<b>August 2012 Changes to the Model Rules Related to Technology .....</b>	<b>29</b>
<b>The Ethics of Computer-Assisted Review: Analysis of the Introductory Hypothetical .....</b>	<b>31</b>
<i>Competence</i> .....	32
<i>Diligence</i> .....	33
<i>Expediting Litigation</i> .....	35
<i>Communication</i> .....	35
<i>Supervision</i> .....	36
<i>Duty of Candor Toward the Tribunal</i> .....	38
<i>Duty of Confidentiality</i> .....	39

<i>Fairness to Opposing Party and Counsel</i> .....	40
<i>Reporting Professional Misconduct</i> .....	40
<i>Potential Litigation Concern: Certification</i> .....	41
<i>Potential Litigation Concern: Inherent Authority of a Court to Impose Sanctions</i> .....	42
<i>Potential Litigation Concern: The Duty to Monitor Production</i> .....	42
<i>Potential Litigation Concern: Sanction for Violating a Rule 16(f) Order</i> .....	42
<b>Conclusion</b> .....	<b>43</b>
<b>Appendix I: Reported Cases Where Judges Have Endorsed the Cooperation Proclamation</b> .....	<b>45</b>
<b>Appendix II: Excerpts From <i>Zubulake V</i>, 229 F.R.D. 422, 432-434 (S.D.N.Y. 2004) on Outside Counsel’s Duty to Monitor Production</b> .....	<b>48</b>
<b>About the Author</b> .....	<b>50</b>
John M. Barkett .....	50

# More on the Ethics of E-Discovery: Predictive Coding and Other Forms of Computer-Assisted Review

**John M. Barkett**  
**Shook, Hardy & Bacon L.L.P.**  
**Miami, Florida**

## **INTRODUCTION - AN INCREASINGLY REALISTIC HYPOTHETICAL**

The cost of e-discovery historically has been in part a function of the time it takes humans to review large numbers of electronic documents reduced to an image on a computer screen associated with a template that is completed by the reviewer to provide information about the document. Large numbers here are not hundreds or thousands of documents, but hundreds of thousands and millions of documents. Imagine a litigation world where, through the use of computer-search technology “trained” by lawyers, linguists or experts, the pool of documents requiring review can be reduced by 50% or 75% or even 90% because a computer has successfully predicted or determined which documents are most likely to be responsive to a production request and only those documents are reviewed. In cases with sufficiently large document volumes or where other measures of cost and benefit justify an effort to reduce the pool of documents needing review, the ability of technology to produce a defensible production by segregating from a large pool of documents a smaller pool of documents that contains either the most responsive documents or most of the responsive documents is information that lawyers must, at a minimum, monitor and, one day soon, may have to embrace.

Consider the following hypothetical.

Your client has been sued. Your case has been assigned to a district court judge and a magistrate judge who have both endorsed the Cooperation Proclamation published by the Sedona Conference. You had already met with your client and the client team to discuss search protocols and techniques with respect to electronically stored information to prepare your case as well as to prepare for production in anticipation of expected requests for electronically stored information. You contact the client’s general counsel to discuss the meet-and-confer session. You tell the general counsel that the issue of search protocols is going to come up. You expect your opponent to raise the issue of keyword search terms. The general counsel tells you that search terms are work product and to tell opposing counsel to issue document production requests and an appropriate response will be made. She then leaves to rush to a meeting.

At the Rule 26 meet-and-confer session, opposing counsel surprises you. She proposes to produce documents resulting from a search for electronic documents based on predictive coding, instead of using an approach

based solely on a keyword search, and seeks your agreement. She refers to the results of TREC Legal Track and the decision of *Da Silva Moore v. Publicis Groupe & MSL Group*, 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Feb. 24, 2012) to legitimize her client's proposed search approach.

You have never heard of TREC Legal Track but don't want to admit it. You have never heard of the *Da Silva Moore* case either but despite your ignorance, you stay cool. You dig in your heels and refuse to agree. Opposing counsel threatens to seek from the court approval to recover the costs of human review if keywords solely are used to identify potentially responsive documents. You stand your ground. You say nothing to your client about this part of the Rule 26 conference.

At the Rule 16 conference, the district court hears from both sides and scolds you for refusing to cooperate with your opponent. The district court gives your opponent the option to search for and produce electronically stored information based solely on predictive coding. Fearing more wrath from the judge, you say nothing further.

As discovery proceeds, you bone up on computer-assisted review and learn a lot more about how to use computer search tools to find relevant documents in a lawsuit. You start badgering opposing counsel with demands for more information about the sample set of documents being used to code the issues in the case and how opposing counsel is making relevance determinations. In the meantime, to fend you off, opposing counsel is listening to the advice of her e-discovery vendor and parroting whatever she is being told. A battle is looming over the size of the sample, the representativeness of the sample documents, coding decisions made by opposing counsel or her vendor (it is not clear who is calling the shots), and your demands for transparency.

What ethical issues might accompany these facts?

Before answering this question, this paper explores the jurisprudence of "cooperation" and "search technology" including the perilous waters lawyers will encounter if they do not comprehend the pitfalls that can accompany e-discovery stipulations. Then it will review the results of the National Institute of Standards and Technology's Text Retrieval Conference "Legal Track," which for the past four years has invited participants to test search technologies on mock production requests from known document sets in the hopes of finding ways to improve information retrieval at a lower cost than might be provided by manual review, yet as effectively or more effectively. That discussion will be followed by a review of two matters. One is *Da Silva Moore*, a case that explains what predictive coding is and that has created e-discovery news headlines for being the first reported decision to have a meaningful discussion of this form of computer-assisted review.<sup>1</sup> The other is *Kleen Products LLC v. Packaging Corp. of America, Inc.*, Civ. No. 1-10-cv-05711 (N.D. Ill.) where the court followed a very different path highlighting the difference between cooperation and capitulation. As will be explained, the court changed the dialogue from plaintiffs' demand that the court order defendants to engage in predictive coding instead of a keyword-based search, to a defendant's right to select the means it wishes to use to make a production. The paper then discusses upcoming "technology" amendments to the Model Rules of Professional Conduct. With this extensive backdrop, the stage will have been set for an analysis of ethical issues associated with this hypothetical.

---

<sup>1</sup> I use the phrase "computer-assisted" review rather than "technology-assisted" review, but these phrases are used interchangeably.

## CASE LAW CALLING FOR COOPERATION

The case law contains unmistakable signals to the Bar on the importance of cooperation in discovery and especially e-discovery.

In a case resulting in sanctions for spoliation of documents after a complaint was filed, *Board of Regents of University of Nebraska v. BASF Corp.* 2007 WL 3342423, at \*5 (D. Neb. Nov. 5, 2007), the magistrate judge cataloged the Civil Rules Advisory Committee's efforts in the 1993, 2000, and 2006 amendments to the Federal Rules of Civil Procedure to increase voluntary production of relevant documents, develop mutually agreeable discovery plans, and require lawyers to confer before a sanctions motion can be brought:

*The overriding theme of recent amendments to the discovery rules has been open and forthright sharing of information by all parties to a case with the aim of expediting case progress, minimizing burden and expense, and removing contentiousness as much as practicable. See, e.g., 1993, 2000, 2006 Amendments to Federal Rules of Civil Procedure, Advisory Committee Comments to Rules 26, 33, 34, 37, 28 U.S.C. Compliance with these changes has placed - on counsel - the affirmative duties to work with clients to make required disclosures, Rule 26(a)(1)(2) and (3); reduce oppression and burden, Rule 26(b)(2); cooperatively plan discovery with opposing counsel, Rule 26(f); affirmatively certify accuracy and good faith in requesting and responding to discovery, Rule 26(g); and confer with opposing counsel to resolve disputes before filing certain motions, Rule 37(a)(2)(B), among others. If counsel fail in this responsibility - willfully or not - these principles of an open discovery process are undermined, coextensively inhibiting the courts' ability to objectively resolve their clients' disputes and the credibility of its resolution.<sup>2</sup>*

A year later, recognizing that e-discovery has placed greater demands on lawyers to cooperate in order to control litigation costs, the Sedona Conference issued a "Cooperation Proclamation" with endorsements by many federal and state court judges.<sup>3</sup> Judges have felt a need to cite or require lawyers to read the Cooperation Proclamation in a growing chorus of reported discovery and e-discovery cases. *See, e.g., Da Silva Moore v.*

---

<sup>2</sup> See also Barkett, *From Canons to Cannon*, in *A Century of Legal Ethics: Trial Lawyers and the ABA Canons of Professional Ethics* (American Bar Association, Chicago, 2009) (contrasting the aspirations contained in the Canons of Ethics with the decline in professionalism and civility among lawyers and the corresponding changes in the civil rules designed to foster more disclosure, more communications among counsel, greater involvement of the court, and better control over discovery).

<sup>3</sup> [http://www.thesedonaconference.org/content/tsc\\_cooperation\\_proclamation/proclamation.pdf](http://www.thesedonaconference.org/content/tsc_cooperation_proclamation/proclamation.pdf) (July 2008). The Proclamation asks lawyers to focus on problem solving not problem expanding when it comes to discovery: "Lawyers preparing cases for trial need to focus on the full cost of their efforts—temporal, monetary, and human. Indeed, all stakeholders in the system—judges, lawyers, clients, and the general public—have an interest in establishing a culture of cooperation in the discovery process. Over-contentious discovery is a cost that has outstripped any advantage in the face of ESI and the data deluge. It is not in anyone's interest to waste resources on unnecessary disputes, and the legal system is strained by 'gamesmanship' or 'hiding the ball,' to no practical effect." Specifically with respect to e-discovery, the Proclamation contains these suggestions: "1. Utilizing internal ESI discovery 'point persons' to assist counsel in preparing requests and responses; 2. Exchanging information on relevant data sources, including those not being searched, or scheduling early disclosures on the topic of Electronically Stored Information; 3. Jointly developing automated search and retrieval methodologies to cull relevant information; 4. Promoting early identification of form or forms of production; 5. Developing case-long discovery budgets based on proportionality principles; and 6. Considering court-appointed experts, volunteer mediators, or formal ADR programs to resolve discovery disputes."

*Publicis Groupe & MSL Group*, 2012 U.S. Dist. LEXIS 23350, \*8 (S.D.N.Y. Feb. 24, 2012); *In re Facebook PPC Advertising Litigation*, 2011 WL 1324516, \*1-2 (N.D. Cal. Apr. 6, 2011).<sup>4</sup>

The Seventh Circuit Pilot Project<sup>5</sup> that began in 2009 contains a number of principles designed to foster coordination with the court; communication with clients, the court, and counsel; and, pertinent here, cooperation. Specifically, Principle 1.02 provides:

*An attorney's zealous representation of a client is not compromised by conducting discovery in a cooperative manner. The failure of counsel or the parties to litigation to cooperate in facilitating and reasonably limiting discovery requests and responses raises litigation costs and contributes to the risk of sanctions.*<sup>6</sup>

The Civil Rules Advisory Committee has taken notice of this activity.

## **A DUTY TO COOPERATE IN RULE 1?**

One theme at the Civil Rules Advisory Committee's "Duke Conference" held in May 2010 at Duke University's Law School was the importance of cooperation to reduce the costs of litigation. As a follow up to the Duke Conference, the "Duke Subcommittee" of the Advisory Committee is considering a proposal to modify Rule 1 of the Federal Rules of Civil Procedure to add a duty to cooperate. Illustrations of possible changes to Rule 1 were included in materials reviewed by the Advisory Committee at its March 22-23, 2012 meeting. The subcommittee first explained that it was wrestling with the ethics of cooperation and advocacy and sought to avoid the history that accompanied Rule 11 after it was first adopted:

*It is easy enough to draft a rule that mandates reasonable cooperation within a framework that remains appropriately adversarial. It is difficult to know whether any such rule can be more than aspirational. Rule 11 already governs unreasonable motion practice, and there is little outcry for changing the standards defined by Rule 11. And there is always the risk that the ploy of adding an open-ended duty to cooperate will invite its own defeat by encouraging tactical motions, repeating the sorry history of the 1983 Rule 11 amendments.*

Addendum to Agenda Materials, Advisory Committee on Civil Rules March 22-23, 2012), p. 37 (footnote omitted). It then offered up proposals that had been presented to it:

---

<sup>4</sup> Appendix I lists additional reported decisions where the Cooperation Proclamation has been endorsed by federal judges. State courts are echoing these federal judges. See, e.g., *In re Weekley Homes, L.P.*, 2009 Tex. LEXIS 630, \*32 (Tex. 2009) ("A fundamental tenet of our discovery rules is cooperation between parties and their counsel, and the expectation that agreements will be made as reasonably necessary for efficient disposition of the case"); "The Texas Lawyer's Creed – A Mandate for Professionalism," issued by the Supreme Court of Texas and the Texas Court of Criminal Appeals (November 7, 1989) (<http://www.supreme.courts.state.tx.us/pdf/TexasLawyersCreed.pdf>). The Texas Lawyer's Creed contains a preamble followed by a number of general and then personal mandates that lawyers in Texas make. One of the general mandates is: "A lawyer owes to opposing counsel, in the conduct of legal transactions and the pursuit of litigation, courtesy, candor, cooperation, and scrupulous observance of all agreements and mutual understandings. Ill feelings between clients shall not influence a lawyer's conduct, attitude, or demeanor toward opposing counsel. A lawyer shall not engage in unprofessional conduct in retaliation against other unprofessional conduct." One of the personal mandates reads: "I will refrain from excessive and abusive discovery."

<sup>5</sup> The Pilot Project is described on the website for the Project at <http://www.discoverypilot.com/>.

<sup>6</sup> For an in-depth discussion of the 7<sup>th</sup> Circuit Pilot Project, see Barkett, *The 7th Circuit E-Discovery Pilot Project: What We Might Learn and Why It Matters to Every Litigant in America* (ABA Section of Litigation News, December 11, 2011 (available at: [http://apps.americanbar.org/litigation/litigationnews/civil\\_procedure/docs/barkett.december11.pdf](http://apps.americanbar.org/litigation/litigationnews/civil_procedure/docs/barkett.december11.pdf)).

*[These rules] should be construed, ~~and~~ administered, and employed by the court and parties to secure the just, speedy, and inexpensive determination of every action and proceeding[, and the parties should cooperate to achieve these ends]*

or

*[These rules] should be construed and administered by the court to achieve the just, speedy, and inexpensive determination of every action and proceeding. The parties should cooperate to achieve these ends.*

No formal amendment is yet before the Advisory Committee for consideration, but the importance of cooperation to bring down the costs of e-discovery has received the imprimatur of too many judges to be ignored.

## **CASE LAW ON COOPERATION AND SEARCH TECHNOLOGY**

The lack of cooperation has been discussed prominently in the case law on protocols for searching electronically stored information. Again, this case law contains unmistakable signals to the Bar.

In *Securities and Exchange Commission v. Collins & Aikman*, 256 F.R.D. 403 (S.D.N.Y. 2009), Judge Scheindlin was blunt in her criticism of the SEC's refusal "to negotiate a workable search protocol" responsive to the requests for production. She called the SEC's position "patently unreasonable" and directed counsel to meet and confer to develop a workable search protocol or a Special Master would be appointed to assist them in doing so. 256 F.R.D. at 415.

In deciding to sanction the producing party's conduct, the magistrate judge in *In re Seroquel Products Liability Litig.*, 244 F.R.D. 650, 662 (M.D. Fla. 2007), was highly critical of the defendant's (AZ) unilateral search efforts:

*[W]hile key word searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process. Rather than working with Plaintiffs from the outset to reach agreement on appropriate and comprehensive search terms and methods, AZ undertook the task in secret...*

*In this case, AZ never discussed with Plaintiffs which search terms to use as part of the search. There was no dialogue to discuss the search terms, as required by Rules 26 and 34. AZ eventually disclosed in April 2007 that a key word search had been conducted, not in seeking collaboration on the words to use, but rather as part of the dialogue on certifying the "custodial" production...*

*AZ has not been as cooperative as possible in resolving the custodial issues. It is undisputed that the production "completed" on June 30, 2007 had load file, metadata, page break and key word search problems, making the 10 million pages of documents inaccessible, unsearchable, and unusable as contemplated under the Rules.*

*United States v. O'Keefe*, 537 F. Supp. 2d 13 (D.D.C. 2008) involved a challenge to a government agency's compliance with court-ordered production of electronically stored information that might aid criminal defendants in their defense. Defendants questioned the government lawyers' selection of "keywords" arguing

that they did not result in the location of documents being sought. Magistrate Judge Facciola, who has made significant contributions to the e-discovery jurisprudence over the years, was not impressed by the lawyers' protestations. He explained that whether keyword searches will yield the information sought "is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics." *Id.* at 24. Hence, if a lawyer is going to claim that a "certain search term or terms" would be more effective than the terms "that were used" is "truly to go where angels fear to tread." He held that a lay person does not have the expert credentials to advance the argument. "Accordingly, if defendants are going to contend that the search terms used by the government were insufficient, they will have to specifically so contend in a motion to compel and their contention must be based on evidence that meets the requirements of Rule 702 of the Federal Rules of Evidence." *Id.*<sup>7</sup>

Magistrate Judge Paul Grimm, who has authored several prominent opinions in the e-discovery arena, has also addressed search protocols. *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008). The context was unusual. Defendants had produced text-searchable electronic files and nontext-searchable electronic files. Among them were privileged documents. Defendants explained that they had done a keyword search of the documents produced and that the search should have resulted in the identification of privileged documents. Because it did not, defendants suggested that the privileged documents were among the nontext-searchable files and their production was inadvertent. Plaintiff disputed these claims and refused to return the privileged documents, claiming waiver. Judge Grimm was not impressed by defendants' keyword search,<sup>8</sup> and citing *O'Keefe*, he rejected defendants' inadvertence claim:

*Defendants, who bear the burden of proving that their conduct was reasonable for purposes of assessing whether they waived attorney-client privilege by producing the 165 documents to the Plaintiff, have failed to provide the court with information regarding: the keywords used; the rationale for their selection; the qualifications of M. Pappas and his attorneys to design an effective and reliable search and information retrieval method; whether the search was a simple keyword search, or a more sophisticated one, such as one employing Boolean proximity operators; or whether they analyzed the results of the search to assess its reliability, appropriateness for the task, and the quality of its implementation. While keyword searches*

---

<sup>7</sup> In *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331 (D.D.C. 2008), Judge Facciola had the opportunity to apply this principle. There was agreement by the parties that a forensic examination of defendant's Macintosh should occur, but the parties were unable to agree on a search protocol. Plaintiff objected to the use of search terms because Lundin had installed a new operating system on his laptop computer and may have thereby compromised the integrity of files previously stored on the computer, plaintiff argued. *Id.* at 333. Plaintiff also argued that a file extension (such as .doc or .jpg) and keyword search would not capture fragments of information contained on a hard drive. The court required plaintiff to submit an affidavit from an expert supporting plaintiff's allegations before deciding how to proceed. *Id.*

<sup>8</sup> "While it is known that M. Pappas (a party) and Mohr and Schmid (attorneys) selected the keywords, nothing is known from the affidavits provided to the court regarding their qualifications for designing a search and information retrieval strategy that could be expected to produce an effective and reliable privilege review. As will be discussed, while it is universally acknowledged that keyword searches are useful tools for search and retrieval of ESI, all keyword searches are not created equal; and there is a growing body of literature that highlights the risks associated with conducting an unreliable or inadequate keyword search or relying exclusively on such searches for privilege review. Additionally, the Defendants do not assert that any sampling was done of the text searchable ESI files that were determined not to contain privileged information on the basis of the keyword search to see if the search results were reliable. Common sense suggests that even a properly designed and executed keyword search may prove to be over-inclusive or under-inclusive, resulting in the identification of documents as privileged which are not, and non-privileged which, in fact, are. The only prudent way to test the reliability of the keyword search is to perform some appropriate sampling of the documents determined to be privileged and those determined not to be in order to arrive at a comfort level that the categories are neither over-inclusive nor under-inclusive. There is no evidence on the record that the Defendants did so in this case." 250 F.R.D. at 256-57.

*have long been recognized as appropriate and helpful for ESI search and retrieval, there are well-known limitations and risks associated with them, and proper selection and implementation obviously involves technical, if not scientific knowledge.*

*Id.* at 259-60 (footnote omitted).

Recognizing the concern some might have about the costs of litigation if experts had to be involved in determining keyword searches, Judge Grimm offered this advice: “For those understandably concerned about keeping discovery costs within reasonable bounds, it is worth repeating that the cost-benefit balancing factors of Fed.R.Civ.P. 26(b)(2)(C) apply to all aspects of discovery, and parties worried about the cost of employing properly designed search and information retrieval methods have an incentive to keep the costs of this phase of discovery as low as possible, including attempting to confer with their opposing party in an effort to identify a mutually agreeable search and retrieval method. This minimizes cost because if the method is approved, there will be no dispute resolving its sufficiency, and doing it right the first time is always cheaper than doing it over if ordered to do so by the court.” *Id.* at 261, n.10.<sup>9</sup>

The magistrate judge in *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, *supra*, was very unhappy with counsel who could not agree on keywords for a search. “This case is just the latest example of lawyers designing keyword searches in the dark, by the seat of the pants, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails.” The lawyer’s impasse left the court “in the uncomfortable position of having to craft a keyword search methodology for the parties, without adequate information from the parties” or the nonparty. The court did so, but he also issued this warning to the Bar:

*Electronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI. Moreover, where counsel are using keyword searches for retrieval of ESI, they at a minimum must carefully craft the appropriate keywords, with input from the ESI’s custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of “false positives.” It is time that the Bar--even those lawyers who did not come of age in the computer era--understand this.*

2009 U.S. Dist. LEXIS 22903, at \*13-14. See also *In re DirecTech Southwest, Inc., Fair Labor Standards Act Litigation*, 2009 U.S. Dist. LEXIS 69142 (E.D. La. Aug. 7, 2009) (in another case of lawyer impasse over keywords, the magistrate judge chose plaintiffs’ proposed keywords rejecting defendants’ argument that plaintiffs’ list would produce many false hits and cost \$100,000 primarily because of the need to conduct a privilege review); *Capitol Records, Inc. v. MP3tunes, LLC*, 2009 WL 2568431, \*50 (S.D.N.Y. Aug. 13, 2009) (“Had the parties focused their attention on discussing their differences, rather than drafting dueling epistles for submission to the Court, MP3tunes undoubtedly would have realized that the EMI Labels were not asking

---

<sup>9</sup> Judge Grimm also noted that a court-approved clawback agreement would have protected defendants, but they failed to seek one. 250 F.R.D. at 261, n.10. Since this decision was rendered Fed. R. Evid. 502 has been adopted. Rule 502(d) allows a litigant to obtain a federal district court order to protect the attorney-client privilege or work product protection associated with an inadvertently produced document and thereby to preserve the privilege or protection in any other federal or state proceeding. Specifically, Rule 502(d) provides in full: “A federal court may order that the privilege or protection is not waived by disclosure connected with the litigation pending before the court – in which event the disclosure is also not a waiver in any other federal or state proceeding.”

to be provided with a transaction history for each search term, so long as they received assurances that the search methodology that MP3tunes employed would lead to the production of all responsive emails.”)

## THE PERILS OF SEARCH IGNORANCE

In addition to the search cooperation cases, other cases comfortably make the point that a lack of understanding of search technologies and costs can haunt a party.

*In re Fannie Mae Securities Litigation*, 552 F.3d 814, 818 (D.C. Cir. 2009) is the poster child. In this case, the Office of Federal Housing Enterprise Oversight, a nonparty, hired 50 contract lawyers to comply with a stipulated agreement to produce documents. The stipulated agreement had given the requesting party the sole right to determine search terms and, at the same time, had waived any claim the nonparty might have had for cost-shifting under Rule 45. The resulting search terms, which OFHEO characterized “as tantamount to a request for the dictionary,” resulted in the retrieval of about 80% of all of the emails of the OFHEO.<sup>10</sup> The OFHEO incurred \$6 million in costs, representing 9% of its annual budget. The nonparty was sanctioned and held in contempt of court for failing to comply with a production deadline in the stipulated agreement. The court of appeals affirmed the contempt finding, refusing to relieve the nonparty from the stipulation. 552 F.3d at 821-22.

In *I-Med Pharma Inc. v. Biomatrix, Inc. et al.*, 2011 U.S. Dist. LEXIS 141614 (D.N.J. Dec. 9, 2011), the district court affirmed a magistrate judge’s order relieving a plaintiff from a stipulation. Plaintiff had agreed to a forensic examination of its computer system. Defendants hired an expert to conduct the forensic examination and apparently selected the search terms, which included these words: “profit,” “loss” “revenue.” The forensic examination included the unallocated space on the plaintiff’s computer system. Because of the “broad search terms,” the number of resulting “hits” came as “no surprise” to the district court: in the unallocated space alone, the terms generated 64,382,929 hits representing an estimated 95 million pages of data. Plaintiff objected to conducting a privilege review of these documents as required by the stipulation and asked the magistrate judge for relief. The judge then entered an order permitting plaintiff to withhold data found in the unallocated space.<sup>11</sup> The judge reasoned that (1) the burden on the plaintiff of reviewing the data from the unallocated space outweighed any potential benefits to the defendants; (2) the complaining defendant had not demonstrated the “complete relevancy” of the information sought and also had not demonstrated that any information had been actually deleted by plaintiff (i.e., there was no justification to review the data from the unallocated space); and (3) the overbroad search terms made the likelihood of finding relevant information that would be admissible at trial “minimal.” *Id.* at \*6. The district court affirmed the order, rejecting the argument

---

<sup>10</sup> The court of appeals was insensitive to the argument that the search term list showed bad faith by the subpoena issuer: “OFHEO insists that the individual defendants’ list of search terms was tantamount to a request for the dictionary, resulting as it did in the retrieval of approximately 80 percent of the office’s emails. (Record citation omitted.) But far from showing bad faith, that figure may simply indicate that most of the emails actually bear some relevance, or at least include language captured by reasonable search terms. More fundamentally, OFHEO does not argue that the individual defendants exercised their contractual rights in bad faith; it argues only that they violated a textual limitation on those rights. As described above, however, that limitation appears nowhere in the stipulated order.” 552 F.3d at 821.

<sup>11</sup> Mindful of the costs that the defendants incurred to extract and search the data in the unallocated space, the judge allow defendants to seek reimbursement of those costs from the plaintiff. 2011 U.S. Dist. LEXIS 141614 at \*6.

that manifest injustice had to be shown to provide relief from a stipulation.<sup>12</sup> The district court added: “While Plaintiff should have known better than to agree to the search terms used here, the interests of justice and basic fairness are little served by forcing Plaintiff to undertake an enormously expensive privilege review of material that is unlikely to contain non-duplicative evidence.” *Id.* at \*17-18.

*Custom Hardware Eng. & Consulting, Inc. v. Dowell et al.*, 2012 U.S. Dist. LEXIS 146 (E.D. Mo. Jan. 3, 2012) involved compliance with a noncompete clause, an agreed-upon forensic examination of defendant’s computer storage media, and a dispute over the breadth of search terms. Plaintiff had broad terms; defendants proposed narrower terms. The district court approved of plaintiff’s search terms because (1) defendants did not provide evidence that these search terms would produce an unreasonable number of search hits; (2) defendants could conduct a privilege review before production was made; and, (3) in the district court’s review it was enough that the search terms “may produce” a document that was relevant in order to approve them. *Id.* at \*9-13. One has to wonder if there will be a follow-up order as occurred in *I-Med* because of the volume of documents generated by the search terms.

*In re National Association of Music Merchants, Musical Instruments and Equipment Antitrust Litigation*, 2011 U.S. Dist. LEXIS 145804 (S.D. Cal. Dec. 19, 2011) involved a cooperative effort on search terms where, after the fact, plaintiff realized that certain abbreviations and acronyms were used by defendants’ employees but had not been part of the search terms utilized. Plaintiffs asked defendants to rerun searches for these abbreviations and acronyms. Defendant refused because plaintiffs had ample opportunity to raise the issue before the search was conducted. The magistrate judge agreed with defendants, explaining that plaintiffs met with defendants before the initial production and the issue had not been raised then, and that the issue was not raised during negotiations between the parties that resulted in an agreement on expanded search terms. The judge also relied on the fact that defendants would have had to review “tens of thousands of additional documents that would likely yield only a very small number of additional responsive documents” if a new search was run. *Id.* at \*22-24.<sup>13</sup>

These cases arose in the context of the relatively simple world of search-term technology. But search technology has advanced beyond mere keywords.

## THE LEGAL TRACK AT TREC

One of the ways to determine the relative worth of different search technologies is to measure their effectiveness in a scientifically controlled evaluation. That’s the goal of the Text Retrieval Conference or TREC.

---

<sup>12</sup> The district court explained: “Clearly a court has the power to modify stipulations concerning discovery terms and deadlines while discovery is still ongoing without the showing of manifest injustice. A court could not effectively perform its duty to fairly and efficiently manage discovery if every minor change to a stipulated briefing schedule or deposition date required a showing of “exceptional circumstances” or “substantial and real harm.” While courts should not casually discard agreements between the parties, nor should they abrogate their duty to balance both burden and the likelihood of uncovering relevant evidence merely because a party made an improvident agreement.” 2011 U.S. Dist. LEXIS 141614 at \*13-14.

<sup>13</sup> The judge, however, did permit the plaintiffs to search for the abbreviations and acronyms in three custodians’ files and one abbreviation in the file of a fourth custodian if plaintiff wanted to bear the search and review expense. 2011 U.S. Dist. LEXIS 145804 at \*\*24.

The National Institute of Standards and Technology (part of the Department of Commerce) has sponsored the TREC Legal Track since 2006. The “Overview of the TREC 2008 Legal Track”<sup>14</sup> (2008 Track Overview) explains its purpose:<sup>15</sup>

*The use of information retrieval techniques in law has traditionally focused on providing access to legislation, regulations, and judicial decisions. Searching business records for information pertinent to a case (or “discovery”) has also been important, but searching records in electronic form was until recently the exception rather than the norm. The goal of the Legal Track at the Text Retrieval Conference (TREC) is to assess the ability of information retrieval technology to meet the needs of the legal community for tools to help with retrieval of business records, an issue of increasing importance given the vast amount of information stored in electronic form to which access is increasingly desired in the context of current litigation. Ideally, the results of a study of how well comparative search methodologies perform when tasked to execute types of queries that arise in real litigation will serve to better educate the legal community on the feasibility of automated retrieval as well as its limitations.*

Magistrate Judge Paul Grimm recognized TREC and the potential it represents to change the way lawyers conduct discovery:

*[T]here is room for optimism that as search and information retrieval methodologies are studied and tested, this will result in identifying those that are most effective and least expensive to employ for a variety of ESI discovery tasks. Such a study has been underway since 2006, when the National Institute of Standards and Technology (NIST), an agency within the U.S. Department of Commerce, embarked on a cooperative endeavor with the Department of Defense to evaluate the effectiveness of a variety of search methodologies. This project, known as the Text Retrieval Conference (TREC), evolved into the TREC LegalTrack, a research effort aimed at studying the e-discovery review process to evaluate the effectiveness of a wide array of search methodologies.*

*Victor Stanley, Inc. v. Creative Pipe, Inc.*, supra, 250 F.R.D. at 261, n.10.<sup>16</sup>

The 2008 Legal Track Overview describes the results of the 2008 Legal Track “Interactive Task,” which was designed to replicate a response to a document request in litigation. The document collection consisted of just under seven million documents drawn from a public database of documents used in tobacco litigation. The designers of the task created a hypothetical complaint and associated document requests (called “topics”). Attorneys were responsible for interpreting topics covered by the document requests and participants in the Interactive Task were given the opportunity to interact with these attorneys as if they had been retained as the e-discovery contractor for a law firm.

---

<sup>14</sup> <http://trec-legal.umiacs.umd.edu/>.

<sup>15</sup> <http://trec.nist.gov/pubs/trec17/papers/LEGAL.OVERVIEW08.pdf>.

<sup>16</sup> In explaining how validation of a retrieval methodology could establish the reasonableness of a party’s electronic production, Judge Grimm added that the TREC Legal Track “can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search, whether for privilege review or other purposes.” 250 F.R.D. at 261, n.10.

The TREC Legal Track measures success of a search in terms of “recall” and “precision.” Recall refers to relevant documents retrieved by a search system as a percentage of all the relevant documents in the document set. If there are 100 relevant documents in a data set of 1000 documents and the search technology locates 25 of them, the recall rate is 25% (25 relevant documents/100 relevant documents in the data set).

Precision represents the relevant documents actually retrieved by a search system as a percentage of all of the documents retrieved from the document set. If the search system utilized locates the 100 relevant documents plus another 100 irrelevant documents, its precision is 50% (100 relevant documents/200 documents retrieved).<sup>17</sup>

A number of automated searches (which included a negotiated set of keywords) had been conducted on the document set. The results of these searches were pooled and this pooled set of searches achieved precision and recall scores of 30-40% in searches of the document set on three different topics. One team that participated in the Interactive Task, in order to design its search protocol, took advantage of the opportunity to interact with a lawyer who served as the “Topic Authority” (the lawyer in charge of responding to the request for production in the hypothetical scenario) in the test.<sup>18</sup> That team participated in searches on one of the three topics. Using an iterative linguistic search approach,<sup>19</sup> it was able to achieve a recall score of 62% (of the relevant

---

<sup>17</sup> A summary measure that represents a participant’s score for both recall and precision is called  $F_1$ . The best combination of recall and precision has the highest  $F_1$ . One has to be careful in placing too much reliance on  $F_1$  scores, however. The  $F_1$  score is a good measure when both recall and precision matter equally. In such circumstances, a high  $F_1$  score will indicate that the search process is performing at a high level of both recall and precision while a low  $F_1$  score will indicate that the search process is performing inadequately in either recall or precision, or both. Even in the case of a low  $F_1$ , however, the component recall and precision scores may be important if, with respect to a search objective, recall matters more than precision or vice versa. For example, if a high recall rate is associated with a small subset of a large document pool, low precision (which would produce a low  $F_1$  score) may not be as important as the reduction in cost by reducing the number of documents requiring review. In an experimental study like TREC’s Legal Track, obtaining measures of recall and precision require an evaluation sample that is distinct from the sample used in training the computer-assisted review system. An independent, statistically-valid, and well quality-controlled evaluation sample will be able to detect errors that result both from deficiencies of the technology and deficiencies in the training set. In litigation where the performance of computer-assisted review is a matter of contention, the ultimate authority on issues of relevance is the judge. How involved the judge will have to be will depend upon how agreeable lawyers and their consultants are on how well a computer-assisted review platform is going to perform or is actually performing and whether parties can agree in advance on an independent validation sample to verify the quality of a production.

<sup>18</sup> Each participant was allowed up to 10 hours of interview time under the test rules but interaction also occurred by e-mail.

<sup>19</sup> For a discussion of search technologies generally, see the *Sedona Conference Best Practices Commentary on the Use of Information Retrieval Methods in E-Discovery*, 2007 Sedona Conference Journal 189 (August 2007). Moving beyond keyword searches, an information retrieval technology might “rely on semantic relations between words, and/or which use “thesauri” to capture documents that would be missed in keyword searching.” *Id.* at 202. Illustratively, “documents about Eskimos and igloos are related to Alaska, even if they do not specifically mention the word ‘Alaska.’” *Id.* A search might involve the use of “taxonomies” and “ontologies” applied to electronically stored information to identify documents for production. “Taxonomy is a hierarchical scheme for representing classes and subclasses of concepts.” *Id.* at 221. If one’s search category is “law personnel,” then any document that mentions lawyer, paralegal, legal assistant, “esq.,” or attorney would be picked up in the concept search. *Id.* “An ontology is a more generic species of taxonomy, often including a wider variety of relationship types than are found in the typical taxonomy. An ontology specifies the relevant set of conceptual categories and how they are related to one another.” *Id.* at 222. To illustrate, if your search category includes “attorneys” then “you may also be interested in documents that use words such as ‘lawyer,’ ‘paralegal,’ or ‘Esq.’” *Id.* A search might also rely “on mathematical probabilities that a certain text is associated with a particular conceptual category.” *Id.* at 203. Taxonomies and ontologies rely on “information that linguists collect from the lawyers and witnesses about the key factual issues in the case, the people, organization, and key concepts relating to the business as well as the idiosyncratic communications that might be lurking in documents, files, and emails.” *Id.* “Machine learning tools” depend on mathematical probabilities not linguistics, and as a result “are arguably helpful in addressing cultural biases of taxonomies. *Id.* Clustering is an example of an unsupervised machine learning process (i.e., a process in which the machine groups similar documents together without relying on an iteratively developed training set). Email from Bruce Hedin, Principal Scientist, H5, to John M. Barkett dated July 17, 2012 (on file with the author). “Clustering” groups “together documents with similar content. There are a variety of ways to define similarity, but one way is to count the number of words that overlap between each pair of documents. The more words they have in common, the more likely they are to be about the same thing.” Sedona Information Retrieval Commentary, at 219. Machine learning can help to find “communications in code language and neologisms. For example, if the labor lawyer were searching for evidence that management was targeting neophytes in the union, she might

documents, it found 62% of them) and a precision score of 81% (of the documents retrieved 81% of them were relevant),<sup>20</sup> considerably in excess of the benchmark scores.<sup>21</sup> There were three other participants conducting searches on this same topic. They had much lower recall scores (in the range of 3-16%) than the benchmark scores.<sup>22</sup>

This variability in recall scores means that computer-assisted search technologies and their implementers are not all equal. This fact is confirmed by the “Overview of the TREC Legal Track” for 2009, 2010, and 2011.<sup>23</sup> Readers are cautioned that the test and evaluation methodology from year-to-year are not the same; hence, it is difficult to draw comparisons of one year’s results to another year’s results.<sup>24</sup>

The 2009 “Interactive” task was similar to the 2008 task but related to a different set of documents and certain design changes.<sup>25</sup> The 2010 Interactive task relied on the same set of documents utilized in 2009 but with new mock requests for production and an added privilege review component. Again, there was a Topic Authority who made decisions on relevance. There was an opportunity for each participating team to consult with the Topic Authority as well before running the team’s software on the data set to generate documents that were

---

miss the term “n00b” (a neologism for “newbie”). This technology, used in government intelligence, is particularly apt in helping lawyers find information when they don’t know exactly what to look for. For example, when a lawyer is looking for evidence that key players conspired to violate the labor union laws, she will usually not know the “code words” or expressions the players may have used to disguise their communications.” *Id.* at 203. Readers are advised that the 2007 Sedona Information Retrieval Commentary is currently undergoing revision to keep pace with search technology improvements since 2007. Email from Kenneth J. Withers, Director of Judicial Education, The Sedona Conference, to John M. Barkett dated July 12, 2012 (on file with the author).

<sup>20</sup> If there were a need to do a human review of the relevant documents, one can see that a precision rate of 81% means that reviewers will not have to spend much time on documents that are not relevant.

<sup>21</sup> 2008 Legal Track Overview, Table 15. Because some of the images were not readable, scores were adjusted to account for Optical Character Recognition quality. At an “OCR-accuracy threshold of 95%” in a document (meaning virtually all of the image was searchable), this participant’s recall was 77% and its precision was 83%. *Id.*, Table 19.

<sup>22</sup> 2008 Legal Track Overview, Table 15. The OCR-adjusted scores (at a 95% OCR-accuracy threshold) were 4-24% for recall and 69-80% for precision. *Id.*, Table 19. While the precision rate was high, the very low recall rates mean that most of the relevant documents were not located by these participants’ search techniques.

<sup>23</sup> The 2009 Overview can be found at <http://trec.nist.gov/pubs/trec18/papers/LEGAL09.OVERVIEW.pdf>. The 2010 Overview appears at <http://trec.nist.gov/pubs/trec19/papers/LEGAL10.OVERVIEW.pdf>. The 2011 Overview appears at <http://trec.nist.gov/pubs/trec20/t20.proceedings.html>.

<sup>24</sup> The 2011 Legal Track Overview provides: “It is also inappropriate to compare the results of TREC 2011 with the results of past TREC Legal Track exercises, as the test conditions as well as the particular techniques and tools employed by the participating teams are not directly comparable.” A similar statement does not appear in the 2010 Overview. It is difficult to discern from the published Overviews whether the results for 2010 and 2009 can be compared to one another. The 2010 Overview says that the Interactive task “reprises the TREC 2009 Interactive task,” Legal Track Overview, 2010, at 1, and the same document collection was used. However, there were modifications made in 2010 to the assessment and the adjudication process, and there were three new requests for production. Legal Track Overview, 2010, at 1, 21-22. Hence, direct year-to-year comparisons appear imprudent.

<sup>25</sup> The 2009 Interactive Task document collection came from a production of e-mails by Enron to the Federal Energy Regulatory Commission. Legal Track Overview, 2009, at 4. It “contained 569,034 unique messages. Together, the 569,034 unique messages have 278,757 attachments, bringing the test collection to a total of 847,791 documents (when parent emails and attachments are counted separately). *Id.*, at 5. The document collection had duplicates. One of the 2009 participants explained that there were a number of duplicates in the data set. “The interactive legal task used a new version of the Enron collection with about 800k ‘documents.’ During the course of doing the task, it became apparent that about half the documents were incorrectly processed duplicates of others in the collection (which were correctly processed). The upshot is that there are about 250K unique messages, and about 100K unique attachments, many of which are vacuous.” Gordon V. Cormack & Mona Mojdeh, *Machine Learning for Information Retrieval: TREC 2009 Web, Relevance Feedback and Legal Tracks*, in NIST SPECIAL PUBLICATION: SP 500-278, THE EIGHTEENTH TEXT RETRIEVAL CONFERENCE (TREC 2009) PROCEEDINGS (2009), available at <http://trec.nist.gov/pubs/trec18/papers/uwaterloo-cormack.WEB.RF.LEGAL.pdf>.)

relevant to the topic in question. Manual assessments of each team's document output and categorization (as relevant or not relevant) again occurred. Each team could also appeal decisions of these assessments.

The 2009 Interactive Task results involved seven topics. Each topic had two to five "runs" of the data that were submitted for evaluation by the various participants. The results on one of the topics were regarded as not reliable since very little use was made of the appeals mechanism.<sup>26</sup> Of the 24 total "runs" submitted, five runs over four topics by three participants had recall rates higher than 70% and four of these five runs had precision rates higher than 70%.<sup>27</sup> On the other hand, seven runs from five other participants had recall rates that ranged from 2.9% to 30.5% with corresponding precision rates of 7.7% to 89.5%.<sup>28</sup>

In 2010, with slight variations in the protocol, there were four topics and 22 submitted runs. The results again show variability in performance across different search strategies. On one topic, recall rates were a challenge: of five runs, all were below 27% suggesting perhaps that the views of the Topic Authority on what was relevant were not well-understood by the participants.<sup>29</sup> Of the documents that the software identified as relevant for this topic, the precision rates (what percentage of the documents retrieved were relevant) range from 58% to 87% for three participants and 19% to 30% for two other participants.

---

<sup>26</sup> Legal Track Overview, 2009, at 17-18. On this topic, there were four runs and the recall rates ranged from 0.9% to 7.6%. The authors argued that the appeal mechanism is "heavily" relied upon to correct errors made in the "first-pass assessment process." Because the appeal mechanism was used only lightly by the participants in this topic but resulted in the correction of errors, the authors felt that "many more errors remain uncorrected."

<sup>27</sup> Legal Track Overview, 2009, Table 6. The F<sub>1</sub> scores in 2009 were higher than 50% for all seven topics in 2009, and higher than 75% for the best performer in five of the seven topics. For an analysis of the 2009 results, see Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, Rich. J.L. & Tech., (2011), <http://jolt.richmond.edu/v17i3/article11.pdf>. The authors describe two computer-assisted review approaches that performed well in the 2009 Legal Track. One utilized an interactive methodology to understand issues and the data well enough to be able to establish relevance criteria. Then each document in a dataset is assessed against the relevance criteria and a determination of whether each document is relevant or not is made. This approach results in the development of automated or semi-automated deterministic queries. The query consists of strings of search instructions (for a simple context, imagine a LEXIS or Westlaw search string) that are very precise so that each string may only capture a few documents but all of the strings together will capture the most relevant documents in a collection. *Id.* at 30 (quoting from an email to Maura R. Grossman, Dec. 17, 2010 from a representative of this TREC Legal Track 2009 participant.) The other approach involved machine learning. The participant took the Wumpus search engine (an information retrieval system developed at the University of Waterloo, see <http://www.wumpus-search.org/>) and coupled it with a custom Web interface "that provided document excerpts and permitted assessments to be coded with a single mouse click." For the four topics involved in 2009, about 12,500 documents were retrieved and reviewed at the average rate of about 3 documents per minute. This participant then used the assessments of these documents to train an "on-line active learning system." The learning system produces an estimate of the probability that each document is relevant. Those probability scores then guided review of the documents through a user interface in decreasing order of score. A previously adjudicated document was skipped in this process. Each document was displayed as text and was coded in a single mouseclick as relevant or not relevant. As a check, this participant also examined documents with high scores (the document was relevant) but that were marked "not relevant" by the human reviewers and also documents with low scores that were marked "relevant." For the four topics, this participant reviewed about 50,000 documents over 42 hours, and revisited the interactive search and judging system to "augment or correct the relevance assessments as new information came to light." The "density" of relevant documents was then estimated "as a function of the score assigned to the documents by the active learning system, based on the assessments rendered during the active learning phase." This participant then used this estimate to decide what the optimum cutoff point was to achieve the best combination of recall and precision. "[T]he end result was that a human reviewed every document produced; however, the number of documents reviewed was a small fraction of the entire document population (14,396 of the 836,165 documents were reviewed, on average, per topic)." *Id.* at 31-34 (footnotes omitted) (citing in large part Gordon V. Cormack & Mona Mojdeh, *Machine Learning for Information Retrieval: TREC 2009 Web, Relevance Feedback and Legal Tracks*, in NIST SPECIAL PUBLICATION: SP 500-278, THE EIGHTEENTH TEXT RETRIEVAL CONFERENCE (TREC 2009) PROCEEDINGS (2009) referenced earlier).

<sup>28</sup> Legal Track Overview, 2009, Table 6.

<sup>29</sup> Legal Track Overview, 2010, at 34. This explanation highlights the fact that good technology will not achieve hoped-for results if the implementers of the technology are unable to accurately define what they are searching for.

For a second topic, which by design had very few relevant documents (575 e-mail messages out of a pool of 455,449 messages), six runs were submitted. Recall rates were all below 25% while the precision on what was assessed by the software as relevant ranged from 69 to 73% on the high end, with three runs between 41 and 48% and one run at 2%.

For a third topic which related to lobbying (suggesting that recall rates should be higher), four of six runs had recall rates above 45%, and three of these four runs had precision rates above 57%, with two runs scoring above 50% on both recall and precision.

The fourth topic related to “privilege.” Two of the five runs on this topic had recall rates above 63% (the other three were below 27%). The precision rates for all five runs ranged from 26% (this run had a recall rate of 71%) to 49% (this run had a recall rate of 7%).<sup>30</sup>

The 2011 Legal Track involved the same dataset utilized in 2010. There were 445,449 distinct email messages that contained 230,143 attachments, or a total of 685,592 documents. There were three topics, each representing a distinct request for production.

In the 2011 TREC Legal Track, a sample of the documents for each topic was taken and assessors coded them as responsive or not. The sample of documents was derived in four ways. First, documents regarded as responsive by the coordinators of the Legal Track who selected the 2011 topics were included. Also included were documents that any of the ten participating teams submitted to a Topic Authority for a determination of whether they were responsive. The Topic Authority made a determination of responsiveness of these documents. Many of these documents were also coded by an assessor from a professional review company whose services were provided *pro bono*. If the assessor’s independent assessment differed from the Topic Authority’s, the Topic Authority reviewed the document again and made the final decision. Any document that any team ranked as among the 100 most probably responsive was also included in the sample without further assessment. These three criteria generated 11,612 documents, or about 2% of the total document set.

Another 5,387 documents were added based on a uniform random sample of the remaining documents in the entire dataset. These documents were reviewed by two assessors and if there was a conflict in the responsiveness determination, the Topic Authority resolved the conflict.

This total of 16,999 documents then became the “gold standard” against which the participants’ assessments would be measured.<sup>31</sup> Based on this gold standard, the estimated number of responsive documents for each topic could be estimated: 20,017 (Topic 401); 3,012 (Topic 402); and 1,239 (Topic 403). These were the documents each participant was trying to capture in the most efficient manner.

Based on the mock document requests, “using *ad hoc* methods,”<sup>32</sup> each participant then identified a pool of potentially responsive documents from the document dataset (called a “seed set”) that it assessed as responsive

---

<sup>30</sup> Legal Track Overview, 2010, Table 20. For two topics, the highest  $F_1$  was 27%. For a third topic it was 41% and for the last topic (relating to lobbying), it was 67%. Testing that involves a mock complaint, a Topic Authority, and evaluation of what is relevant is informative but has its limits. In the end there is no substitute for determining costs and benefits in a real case.

<sup>31</sup> Legal Track Overview, 2011, at 2-3.

<sup>32</sup> The “*ad hoc*” methods were not described but involved automated review (using, e.g., keywords or metadata or prior information known about the documents). Maura Grossman, a principal author for the 2011 Overview, provided this example: “A Google search would represent an *ad hoc* review—it finds relevant documents and perhaps the most relevant documents, but it will not necessarily find all relevant documents.” Personal Communication, July 14, 2012.

or nonresponsive. Based on the responsive determinations for the seed set, each participant then ranked the documents in the entire collection of 685,592 documents from “most to least likely to be responsive” linking each document to a probability of responsiveness.<sup>33</sup> A review process could then be conducted based on the ranking—reviewing documents from most responsive to least responsive but stopping at some point in the ranking where it did not appear that the effort to review was going to improve the quality of the production, or what is referred to as the “cutoff” point.<sup>34</sup>

Before submitting their final set of probability (of responsiveness) estimates, each team was permitted to obtain responsiveness determinations from the Topic Authority at three points: up to 100 requests for determination were allowed before an interim set of probability estimates were submitted; up to 200 requests were allowed before a second interim submission; and up to 700 requests were allowed before the final submission.<sup>35</sup> There was also a “mopup” opportunity where each team could review the results of responsiveness determinations received by any other team. The results of these assessments were used by teams to train their computer-assisted review system before producing computer “runs” of the documents that represented each participant’s evaluation of the responsiveness of a document relative to the gold standard.

For 2011, each team had to declare whether each of its runs of the dataset were “automatic” or “technology-assisted.” Automatic runs “were allowed to use manual query formulation, but human review of the document collection was not permitted” (other than that provided in the responsiveness determinations). Participants who utilized technology-assisted runs were allowed to “avail themselves of any amount of human review.”<sup>36</sup> Each team also had to provide information on the number of hours spent to configure the systems used, search for documents, review documents and analyze results prior to submission.<sup>37</sup>

There were three topics, two cast as a new request for production relating to the mock complaint used for the 2009 Legal Track study, and the other as a new request relating to the complaint for the 2010 Legal Track study.

The goal of each participant was to yield as large a population of responsive documents in the upper echelons of the ranking, so that at lower and lower cutoff points, there would still be a high recall rate; i.e., a large number of responsive documents relative to the responsive documents identified by the gold standard. For each of the runs ranking the responsiveness generated by each participant’s search technologies for the 685,592 documents in the dataset, various cutoff points were picked against which recall and precision rates of each

---

<sup>33</sup> “The two learning objectives—ranking and estimating the likelihood of responsiveness—may be accomplished by the same method or by different methods. Either may be automated or manual. For example, ranking may be done using an information retrieval method or by human review using a five-point scale. Estimation may be done in the course of ranking or, for example, by sampling and reviewing documents at representative ranks.” Legal Track Overview, 2011, at 3.

<sup>34</sup> Legal Track Overview, 2011, at 4. The 2011 Overview also explains that one might employ a “triage” strategy by reviewing only mid-ranked documents—picking a cutoff point from the top and the bottom where documents are determined to be responsive or nonresponsive, respectively, and then evaluating the mid-range documents to better establish the final cutoff point for the production. This strategy, however, was not evaluated in the TREC 2011 Legal Track.

<sup>35</sup> Legal Track Overview, 2011, at 4-5.

<sup>36</sup> Legal Track Overview, 2011, at 5.

<sup>37</sup> Legal Track Overview, 2011, Table 3, at 6 (Depending upon the run, the time spent totaled anywhere from 3 hours to 655 hours).

participant were measured. The cutoff points chosen for depiction in tables in the 2011 Overview were 2,000, 5,000, 20,000, 50,000, 100,000, and 200,000 documents for each of the three topics.<sup>38</sup>

The ranges of recall reported are summarized in the table below. The variability in the ranges is stark.<sup>39</sup>

<b>Cutoff Point of Documents Ranked from Most Responsive to Least Responsive</b>	<b>Range of Recall for 28 Runs for Topic 401</b>	<b>Range of Recall for 18 Runs for Topic 402</b>	<b>Range of Recall for 24 Runs for Topic 403</b>
2,000	1% - 10%	2% - 51%	1% - 60%
5,000	3% - 22%	4% - 58%	2% - 95%
20,000	5% - 53%	13% - 75%	3% - 97%
50,000	8% - 70%	15% - 88%	4% - 100%
100,000	8% - 82%	15% - 99%	4% - 100%
200,000	18% - 96%	31% - 100%	4% - 100% <sup>40</sup>

Precision (the percentage of relevant documents among the documents found) tended to be higher for the cutoff points with the smallest number of documents and higher for the cutoff points with the larger number of documents. For Topic 403, no precision rate was higher than 50% for any of the 24 runs at any of the document cutoff levels. For Topic 402, there were two precision rates higher than 50%, one at 57% and the other at 77% and both at the 2,000 document cutoff level, and they were associated with recall rates of 38% and 51% respectively. For Topic 401, at the 2,000 and 5,000 document cutoff levels, there were a number of precision rates above 50%. The highest associated recall rate was 22%—not surprisingly, perhaps, given the low document cutoff point. Two precision rates were higher than 50% for the remaining document cutoff levels (20,000 to 200,000) for all 28 runs, both at 51%; they were associated with recall rates of 50% and 53% respectively.<sup>41</sup>

<sup>38</sup> Legal Track Overview, 2011, Tables 5-7, at 8-9, 11.

<sup>39</sup> This variability is given a bit of a context in the 2011 Overview: “Some participants may have conducted an all-out effort to achieve the best possible results, while others may have conducted experiments to illuminate selected aspects of document review technology. It is inappropriate – and forbidden by the TREC participation agreement – to claim that the results presented here show that one participant’s system or approach is generally better than another’s.... One may see from the results presented in this Overview that some particular techniques and tools achieved good results in this exercise, and therefore show promise that they might also achieve good results in other document review efforts. The efficacy of the participants’ efforts are characterized by the quality of ranking and the accuracy of recall estimates. Efficacy must be interpreted in light of effort, which is characterized by the number of relevance determinations sought from the Topic Authority, as well as by the amount of manual effort employed by the participating team.” Legal Track Overview, 2011, at. 9.

<sup>40</sup> Legal Track Overview, 2011, at 8, 10-11.

<sup>41</sup> Legal Track Overview, 2011, at 8, 10-11. If one’s goal is to reduce the size of a document pool that has to be reviewed as is often the case in litigation, low precision rates are not important if there is a high recall rate associated with a low document cutoff point. One can also see that if one needs both high recall and precision over an entire document collection, the choice of the appropriate computer-assisted review system may require greater probing.

The participants were also asked to estimate the recall rate that they thought they had achieved at every cutoff level. As the 2011 Overview states, “The results are not encouraging. Most runs for most topics dramatically overestimated recall at all cutoff levels. Such an overestimate might lead the manager of a review effort to terminate the review prematurely, due to the false belief that a high level of recall had been achieved.”<sup>42</sup>

On the other hand, some of the submissions made by the participants demonstrated that, depending upon the topic, less than 1%, 3%, and 11%, respectively, of the 685,592 documents had to be reviewed to achieve recall rates of about or in excess of 70%. In other words, 70% of the responsive documents (determined in relation to the “gold standard” explained above) were within the rankings of these submissions for cutoff points of 5,000 documents (Topic 403), 20,000 documents (Topic 402), and 75,000 documents (Topic 401). And because the documents were ranked by “most responsive” to “least responsive,” one can see how production costs could be reduced significantly if these kinds of percentages could be achieved in litigation involving larger document volumes.

The 2011 Overview contains a good summary of the TREC Legal Track evaluation of search technologies as well as the limitations inherent in the studies:

*From 2008 through 2011, the results show that the technology-assisted review efforts of several participants achieve recall scores that are about as high as might reasonably be measured using current evaluation methodologies. These efforts require human review of only a fraction of the entire collection, with the consequence that they are far more cost-effective than manual review. There is still plenty of room for improvement in the efficiency and effectiveness of technology-assisted review efforts, and, in particular, the accuracy of intra-review recall estimation tools, so as to support a reasonable decision that “enough is enough” and to declare the review complete. Commensurate with improvements in review efficiency and effectiveness is the need for improved external evaluation methodologies that address the limitations of those used in the TREC Legal Track and similar efforts. How can we construct a gold standard with reasonable effort, or, in the alternative, measure review effectiveness without a gold standard? How best can we measure recall and precision values*

---

<sup>42</sup> Legal Track Overview, 2011, at 10. Lawyers always believe “smoking guns” or, at least, “hot” documents, exist in a document collection. One would think that they would be captured by a proper computer-assisted review but if, for some reason, they happen to be located in the less responsive or nonrelevant document pool and the cutoff point chosen excludes review of those documents, they would be missed. Sampling the “no review” document pool to verify that they were not missed is one solution. But the possibility that they might be missed will almost certainly be a focal point of contested efforts to use or demand computer-assisted review in a production in litigation. Cf. *Scotts Co., LLC v. Liberty Mutual Insurance Co.*, 2007 WL 1723509, \*6 (S.D. Ohio June 12, 2007) (request to conduct a forensic examination of producing party’s computer systems, network servers, and databases based on the “mere suspicion” that defendant may be withholding discoverable information was not a sufficient basis upon which to grant the request); *Hubbard v. Potter*, 2008 WL 43867, \*4 (D.D.C. Jan. 3, 2008) (plaintiffs’ speculation that more electronic information existed was insufficient: “I cannot find on this record that the theoretical possibility that other electronic documents might exist justifies the additional discovery plaintiffs seek”). *Williams v. Massachusetts Mut. Life. Ins. Co. et al.*, 226 F.R.D. 144, 146 (D. Mass. 2005) (in response to requesting party’s belief that a smoking gun email existed which had not been found and a demand to allow the requesting party to do his own search of the producing party’s data storage systems, the court explained: “Before permitting such an intrusion into an opposing party’s information system--particularly where, as here, that party has undertaken its own search and forensic analysis and has sworn to its accuracy--the inquiring party must present at least some reliable information that the opposing party’s representations are misleading or substantively inaccurate.”) See, *generally*, Barkett, *E-Discovery: Twenty Questions and Answers*, (Chicago: First Chair Press, 2008), at 13-20.

*that are beyond the limit of what can be measured with reference to a single assessor? How can we better control for the amount of effort expended in conducting document review?*<sup>43</sup>

In slightly different words, one can take from TREC's Legal Track two important principles: (1) there can be wide variability in the performance of computer-assisted document review systems—*caveat emptor* are important words to remember if one is considering a computer-assisted review platform; and (2) there is an upfront dollar cost associated with machine learning (like predictive coding) or automated or semi-automated forms of computer assisted review that has to be evaluated in relation to the cost of keyword or other forms of, what are today, conventional search strategies.

TREC's Legal Track also shows that whatever the computer-assisted review platform, judgment—human judgment—is not eliminated. To the contrary, there is a premium placed on exercising good judgments in planning the computer-assisted review. Poor judgment about what is relevant, the necessary size of a sample set, or the documents needed in the sample set to be representative of the document collection, or poor information gathering about the issues in controversy and the linguistics associated with those issues or poor formulations of search strings to determine relevant documents, may produce the same results as a paper production where decisions about responsiveness by the producing party differ from the views of the requesting party and potentially relevant documents were not reviewed at all or were reviewed but not produced.<sup>44</sup>

As this research receives additional judicial scrutiny and as rules of professional responsibility place a greater burden on lawyers to stay abreast of changes in technology, in appropriate matters lawyers are going to have to become more familiar with search methods and protocols beyond keyword searches and with methods to evaluate their effectiveness. Whether they are supporting a particular protocol or opposing it, they will also likely need access to professionals with the expertise to explain the protocols.

This now brings us to *Da Silva Moore* and *Kleen Products*.

## **DA SILVA MOORE**

In *Da Silva Moore*, according to the magistrate judge, the parties agreed to utilize predictive coding but could not agree on the protocol to implement the search technology. In this posture, Judge Peck issued his February 24, 2012 order summarizing his resolution of the parties' differences and offering insights on the difficulties that parties and judges would face in evaluating the propriety of a predictive coding tool in a particular matter.

Relying on an article he had written,<sup>45</sup> Judge Peck described computer-assisted coding as follows:

*Unlike manual review, where the review is done by the most junior staff, computer-assisted coding involves a senior partner (or [small] team) who review and code a "seed set" of documents. The computer identifies properties of those documents that it uses to code other*

---

<sup>43</sup> Legal Track Overview, 2011, at 10, 16

<sup>44</sup> Keyword searches are another form of computer-assisted review. If human judgments result in a faulty design they, too, can be over-inclusive or under-inclusive resulting either in too many documents and potentially costly review or too few of the documents that are relevant.

<sup>45</sup> "Search, Forward," Law Technology News (October 1, 2011).

*documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer's coding. (Or, the computer codes some documents and asks the senior reviewer for feedback.)*

*When the system's predictions and the reviewer's coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents. Typically, the senior lawyer (or team) needs to review only a few thousand documents to train the computer.*

*Some systems produce a simple yes/no as to relevance, while others give a relevance score (say, on a 0 to 100 basis) that counsel can use to prioritize review. For example, a score above 50 may produce 97% of the relevant documents, but constitutes only 20% of the entire document set.*

*Counsel may decide, after sampling and quality control tests, that documents with a score of below 15 are so highly likely to be irrelevant that no further human review is necessary. Counsel can also decide the cost-benefit of manual review of the documents with scores of 15-50.*

2012 U.S. Dist. LEXIS at \*6-7.

Judge Peck then proceeded to address the parties' differences. In making these determinations, the judge kept his eye on what the case was about: whether defendant MSL and its parent company, Publicis, engaged in sex discrimination against five women by limiting their advancement as part of systemic, company-wide gender discrimination. He also separated discovery into two phases allowing the parties the chance to be targeted, and perhaps cautious, in phase one with the opportunity to fill in any discovery gaps in phase two.

### **Custodians**

The court excluded certain emails of "male comparators" from the emails that would be subjected to "predictive coding review" because the search of these excluded emails (intended, plaintiffs argued, to show the job duties of these males and how their duties compared to plaintiffs' duties) "would be so different from that of the other custodians." Judge Peck decided that these other custodians could be deposed to develop information sought with respect to them rather than allowing a separate search of these emails.

The judge also excluded from predictive coding review the emails of the Chief Executive Officer of one of the defendants who worked in France. Judge Peck noted that these emails were "mostly written in French." He reasoned that these emails would be among those recovered from recipients of emails that were included in the predictive coding review dataset. He was also concerned that the emails stored in France were covered by French privacy and blocking statutes.

Other custodians who worked in offices where no named plaintiff worked were also excluded because there was not a certified class and, until then, Judge Peck was limiting discovery to offices where the named plaintiffs worked.

### **Sources of ESI**

The parties debated over inclusion of "shared" folders. The parties could not explain what these folders contained. Hence, Judge Peck left them to phase two unless the parties could better describe their contents.

The judge also explained that there “has to be a limit on redundancy” to comply with the proportionality limits of Rule 26(b)(2)(C). *Id.* at \*15.

### **Predictive Coding Protocol**

The parties had agreed to use a 95% confidence level (plus or minus 2%) to create a random sample of the emails that were going to be searched. To achieve this confidence level of relevant and irrelevant documents, 2,399 documents were identified for the “seed set” to use “to train the predictive coding software.” *Id.* at \*16.<sup>46</sup>

Judge Peck then explained the protocol as to the creation of the seed set:

*To further create the seed set to train the predictive coding software, MSL coded certain documents through “judgmental sampling.” The remainder of the seed set was created by MSL reviewing “keyword” searches with Boolean connectors (such as “training and Da Silva Moore,” or “promotion and Da Silva Moore”) and coding the top fifty hits from those searches. MSL agreed to provide all those documents (except privileged ones) to plaintiffs for plaintiffs to review MSL’s relevance coding. In addition, plaintiffs provided MSL with certain other keywords, and MSL used the same process with plaintiffs’ keywords as with the MSL keywords, reviewing and coding an additional 4,000 documents. All of this review to create the seed set was done by senior attorneys (not paralegals, staff attorneys or junior associates). MSL reconfirmed that “[a]ll of the documents that are reviewed as a function of the seed set, whether [they] are ultimately coded relevant or irrelevant, aside from privilege, will be turned over to” plaintiffs.*

*Id.* at \*16-17 (record citations omitted).

MSL’s vendor used software that ranked documents on a score of 100 (most relevant) to zero (least relevant). MSL proposed to conduct seven iterative rounds of searching. In each round, MSL would review at least 500 documents “from different concept clusters to see if the computer is returning new relevant documents.” “After the seventh round, to determine if the computer is well trained and stable, MSL would review a random sample (of 2,399 documents) from the discards (i.e., documents coded as non-relevant) to make sure the documents determined by the software to not be relevant do not, in fact, contain highly-relevant documents. For each of the seven rounds and the final quality-check random sample, MSL agreed that it would show plaintiffs all the documents it looked at including those deemed not relevant (except for privileged documents).” *Id.* at \*17-18.

---

<sup>46</sup> The size of the seed set is a subject of contention. Plaintiffs objected to Judge Peck’s decision. One of those objections appears in an expert declaration challenging the protocol approved by Judge Peck. With respect to the size of the seed set, the expert opined that a sample size of 16,555, not 2,399, documents was needed to make the computer-assisted review statistically valid. The expert wrote: “Using Defendant MSL’s own numbers, a random sample of only 2,399 documents is not scientifically supportable. Brett Anders, counsel for Defendant MSL, stated at the January 4, 2012 hearing that his review of an initial random sample of documents indicated that the ultimate percentage of responsive documents would be 1.5% of the total population. Based on this 1.5% number, one can predict that, of the approximately 2.5 million documents subject to the ESI protocol, 37,500 documents would be responsive to Plaintiffs’ requests, and thus identified as ‘relevant.’ Conversely, the remaining 2,462,500 documents would be identified as ‘irrelevant’ by the system.” Declaration of Paul J. Neale, Docket Entry 125 Civ. No. 11-CV-1279 (S.D.N.Y. March 19, 2012) (record citation omitted).

Judge Peck accepted MSL's proposal with this caveat: "But if you get to the seventh round and [plaintiffs] are saying that the computer is still doing weird things, it's not stabilized, etc., we need to do another round or two, either you will agree to that or you will both come in with the appropriate QC information and everything else and [may be ordered to] do another round or two or five or 500 or whatever it takes to stabilize the system." *Id.* at \*19.

In response to further objections from plaintiffs about the reliability of MSL's proposed approach, Judge Peck responded as follows:

1. Plaintiff's concerns are premature.
2. He would closely supervise the process.
3. MSL's proposed approach is "totally transparent." Plaintiffs "will see how MSL has coded every email used in the seed set (both relevant and irrelevant)."
4. The Court would be available to resolve any issues.
5. To evaluate proportionality, it is prudent to phase the production: "In order to determine proportionality, it is necessary to have more information than the parties (or the Court) now has, including how many relevant documents will be produced and at what cost to MSL. Will the case remain limited to the named plaintiffs, or will plaintiffs seek and obtain collective action and/or class action certification? In the final sample of documents deemed irrelevant, are any relevant documents found that are 'hot,' 'smoking gun' documents (i.e., highly relevant)? Or are the only relevant documents more of the same thing? One hot document may require the software to be re-trained (or some other search method employed), while several documents that really do not add anything to the case might not matter. These types of questions are better decided 'down the road,' when real information is available to the parties and the Court." *Id.* at \*25-26.<sup>47</sup>

### Lessons for the Future

Judge Peck then discussed published research to establish that manual review of documents and keyword searches are not necessarily the most cost-effective alternative in every case and would not be in this matter:

*Linear manual review is simply too expensive where, as here, there are over three million emails to review. Moreover, while some lawyers still consider manual review to be the "gold standard," that is a myth, as statistics clearly show that computerized searches are at least as accurate, if not more so, than manual review.*

\*\*\*

*Because of the volume of ESI, lawyers frequently have turned to keyword searches to cull email (or other ESI) down to a more manageable volume for further manual review. Keywords have a place in production of ESI -- indeed, the parties here used keyword searches (with Boolean connectors) to find documents for the expanded seed set to train the predictive coding software. In too many cases, however, the way lawyers choose keywords is the*

---

<sup>47</sup> Judge Peck also rejected objections that acceptance of MSL's predictive coding protocol violated Fed. R. Evid. 702 and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). "Rule 702 and Daubert are simply not applicable to how documents are searched for and found in discovery." 2012 U.S. Dist. LEXIS 23350, \*23-24.

*equivalent of the child's game of "Go Fish." The requesting party guesses which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party's "cards" (i.e., the terminology used by the responding party's custodians). Indeed, the responding party's counsel often does not know what is in its own client's "cards."*

*Another problem with keywords is that they often are over-inclusive, that is, they find responsive documents but also large numbers of irrelevant documents. In this case, for example, a keyword search for "training" resulted in 165,208 hits; Da Silva Moore's name resulted in 201,179 hits; "bonus" resulted in 40,756 hits; "compensation" resulted in 55,602 hits; and "diversity" resulted in 38,315 hits. If MSL had to manually review all of the keyword hits, many of which would not be relevant (i.e., would be false positives), it would be quite costly.*

*Moreover, keyword searches usually are not very effective. In 1985, scholars David Blair and M. Maron collected 40,000 documents from a Bay Area Rapid Transit accident, and instructed experienced attorney and paralegal searchers to use keywords and other review techniques to retrieve at least 75% of the documents relevant to 51 document requests. David L. Blair & M. E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, 28 *Comm. ACM* 289 (1985). Searchers believed they met the goals, but their average recall was just 20%. *Id.* This result has been replicated in the TREC Legal Track studies over the past few years.*

*Id.* at \*28, \*31-32 (footnote and record citations omitted).<sup>48</sup>

---

<sup>48</sup> Judge Scheindlin has weighed in on the topic. *National Day Laborer Organizing Network v. United States Immigration and Customs Enforcement Agency*, 2012 U.S. Dist. Lexis 97863 (S.D.N.Y. July 13, 2012) is an action brought under the Freedom of Information Act (FOIA) to obtain records from federal agencies related to a "Secure Communities," an immigration enforcement program started in 2008 that, in part, involves sharing fingerprints of persons arrested by the Federal Bureau of Investigation with the Department of Homeland Security so that they can be checked against immigration records. Plaintiffs' FOIA request resulted in a negotiated agreement called the Rapid Production List (RPL). Defendants ultimately produced "tens of thousands" of responsive records from searches that involved "hundreds of employees and thousands of hours." The parties then cross-moved for summary judgment on whether the defendant agencies met their burden under the FOIA "to conduct search reasonably designed to identify and locate responsive documents." *Id.* at \*8. Citing to a number of decisions, the district court described the agencies' burden under FOIA: (1) identify searched files, (2) describe at least generally the structure of the agency's file system, (3) establish that the search embraced all custodians who were reasonably likely to possess responsive documents, and (4) set forth the search terms used and the type of search performed. *Id.* at \*10 (citations omitted). After addressing arguments over whether certain custodians' files should have been included in the searches, the district court moved to the adequacy of the searches themselves. Explaining that inadequately designed searches will miss responsive documents, the court stated that it could not determine the reasonableness of the searches made by each agency. The court first noted that the affidavits submitted by defendants do "not record and report the search terms that it used, how it combined them, and whether it searched the full text of documents." Then citing to the court's own casebook, an article on alternatives to key word searching, and Judge Peck's decision in *Da Silva Moore*, the district court stated that keyword searching "is often not enough," and may not be effective. *Id.* at \*47. Judge Scheindlin acknowledged that a FOIA response is not held to the same responsiveness standard as a response under the Federal Rules of Civil Procedure, but still regarded the e-discovery case law as applicable: "The failure to actually uncover all documents is not fatal to the adequacy of the government's search; but the failure to design a search that is reasonably calculated to uncover all documents is fatal. Therefore, much of the logic behind the increasingly well-developed caselaw on e-discovery searches is instructive in the FOIA search context because it educates litigants and courts about the types of searches that are or are not likely to uncover all responsive documents." *Id.* at \*48, n.110 (emphasis in original). Relying on a "review of the literature," the district court then added: "[B]eyond the use of keyword search, parties can (and frequently should) rely on latent semantic indexing, statistical probability models, and machine learning tools to find responsive documents. Through iterative learning, these methods (known as "computer-assisted" or "predictive" coding) allow humans to teach computers what documents are and are not responsive to a particular FOIA or discovery request and they can significantly increase the effectiveness and efficiency of searches. In short, a review of the literature makes it abundantly clear that a court cannot simply trust the defendant agencies'

On the other hand, Judge Peck also identified the lessons for future cases that are taught by his experience in this case:

- 1. The results of computer-assisted review will require monitoring and evaluation by the parties and court to provide assurances that it is working as intended:** “[I]t is unlikely that courts will be able to determine or approve a party’s proposal as to when review and production can stop until the computer-assisted review software has been trained and the results are quality control verified. Only at that point can the parties and the Court see where there is a clear drop off from highly relevant to marginally relevant to not likely to be relevant documents. While cost is a factor under Rule 26(b)(2)(C), it cannot be considered in isolation from the results of the predictive coding process and the amount at issue in the litigation.”
- 2. Production in phases makes sense:** “[S]taging of discovery by starting with the most likely to be relevant sources (including custodians), without prejudice to the requesting party seeking more after conclusion of that first stage review, is a way to control discovery costs. If staging requires a longer discovery period, most judges should be willing to grant such an extension.”
- 3. Proactive disclosure of key custodians and the proposed search approach increase the likelihood of support for the search approach:** “[I]n many cases requesting counsel’s client has knowledge of the producing party’s records, either because of an employment relationship as here or because of other dealings between the parties (e.g., contractual or other business relationships). It is surprising that in many cases counsel do not appear to have sought and utilized their client’s knowledge about the opposing party’s custodians and document sources. Similarly, counsel for the producing party often is not sufficiently knowledgeable about their own client’s custodians and business terminology. Another way to phrase cooperation is ‘strategic proactive disclosure of information,’ i.e., if you are knowledgeable about and tell the other side who your key custodians are and how you propose to search for the requested documents, opposing counsel and the Court are more apt to agree to your approach (at least as phase one without prejudice).”
- 4. Involvement of e-discovery vendors can be invaluable:** “[T]he Court found it very helpful that the parties’ ediscovery vendors were present and spoke at the court hearings where the ESI Protocol was discussed.... Even where as here counsel is very familiar with ESI issues, it is very helpful to have the parties’ ediscovery vendors (or in-house IT personnel or in-house ediscovery counsel) present at court conferences where ESI issues are being discussed. It also is important for the vendors and/or knowledgeable counsel to be able to explain complicated ediscovery concepts in ways that make it easily understandable to judges who may not be tech-savvy.”

*Id.* at \*37-40.

#### **The District Court’s Adoption of Judge Peck’s February 24, 2012 Order**

The district court judge, Andrew Carter, adopted Judge Peck’s February 24, 2012 order. *Da Silva Moore v. Publicis Group et al.*, 2012 U.S. Dist. LEXIS 58742 (S.D.N.Y. Apr. 25, 2012).

---

unsupported assertions that their lay custodians have designed and conducted a reasonable search.” *Id.* at \*50 (footnotes omitted). All of this is *dicta*, however. Despite the court’s endorsement of computer-assisted review tools, it did not require their use. Rather, it required the parties to reach agreement on search terms and protocols and “if they wish to and are able to, then they may agree on predictive coding techniques and other more innovative ways to search.” *Id.* at \*57.

Judge Carter rejected plaintiffs' objection that they had not, in fact, acquiesced in the use of predictive coding as Judge Peck had written. The district court explained that the confusion over whether there was an agreement to use the technology was "immaterial" because "the ESI protocol contains standards for measuring the reliability of the process and the protocol builds in levels of participation by Plaintiffs." The district court then elaborated:

*It provides that the search methods will be carefully crafted and tested for quality assurance, with Plaintiffs participating in their implementation. For example, Plaintiffs' counsel may provide keywords and review the documents and the issue coding before the production is made. If there is a concern with the relevance of the culled documents, the parties may raise the issue before Judge Peck before the final production. Further, upon the receipt of the production, if Plaintiffs determine that they are missing relevant documents, they may revisit the issue of whether the software is the best method. At this stage, there is insufficient evidence to conclude that the use of the predictive coding software will deny Plaintiffs access to liberal discovery.*

*Id.* at \*7.<sup>49</sup>

Judge Carter also rejected attacks on the reliability of predictive coding review calling them "premature" and then reassuringly adding: "If the method provided in the protocol does not work or if the sample size is indeed too small to properly apply the technology, the Court will not preclude Plaintiffs from receiving relevant information, but to call the method unreliable at this stage is speculative." *Id.* at \*7-8.

Then, again in language that appears to endorse the selection of predictive coding over other forms of review irrespective of the agreement of the parties, Judge Carter held: "Judge Peck concluded that under the circumstances of this particular case, the use of the predictive coding software as specified in the ESI protocol is more appropriate than keyword searching. The Court does not find a basis to hold that his conclusion is clearly erroneous or contrary to law." *Id.* at \*8-9.

The "jury" thus is now out on how well the predictive coding tool will work in *Da Silva Moore*<sup>50</sup> and whether the protocol and seed set decisions made by Judge Peck will prove to be cost-saving both in terms of the

---

<sup>49</sup> Perhaps Judge Carter was, without saying so, expressing the same sentiment that, as discussed below, ultimately swayed Judge Nolan in *Kleen Products*: a producing party should be entitled to select its production method.

<sup>50</sup> The parties have been sparring over what documents should be included as responsive to populate the seed set. In a letter brief to Judge Peck dated April 27, 2012, plaintiffs' counsel told the court that, "Among the documents that MSL coded as non-relevant are MSL's own policy manual, numerous documents relating to the seven plaintiffs, e-mails showing a centralized team of decision-makers granting exceptions to the salary and hiring freeze imposed by Publicis Groupe, and documents relating to the Company-wide reorganization..." Plaintiffs added: "MSL, however, marked as non-relevant hundreds of documents regarding pay, promotion and other employment decisions that, although "substantially similar" to documents marked as relevant, did not involve the Named Plaintiffs. Defense counsel argued that such documents were properly excluded from the seed set because the Court had limited the scope of class discovery. ...Here, because the coding of the seed set is not just an end in itself, but a means of training the system to locate relevant documents, such coding is guaranteed to confuse the system by indicating that the same concept is both relevant and non-relevant." Docket Entry 193-1, Exhibit PP, Case 1:11-cv-01279 (S.D.N.Y.). In a hearing on May 7, 2012, Judge Peck then reviewed several individual documents to determine whether they were relevant and should be included in the seed set. *Id.*, Exhibit HH. On May 14, 2012, Judge Peck stayed production of electronic discovery pending resolution of plaintiffs' motion for collective action certification and to amend their complaint. Docket Entry 196, Case 1:11-cv-01279 (S.D.N.Y.).

parties and the court's resources, without compromising the administration of justice.<sup>51</sup> But as *Kleen Products* shows, judges are not going to have an easy time figuring out whether to order computer-assisted review by a party who prefers to produce documents in a different manner.

## KLEEN PRODUCTS

There is a fundamental tension between cooperation in discovery and advocating for a client. The judicial calls for cooperation are, in large measure, a plea to solve problems upfront before they turn into sanctions motions. Cooperating on keyword search terms to avoid a fight over the quality of those keyword search terms after a production has already been made is case management: if a judge can preempt motion practice over the defensibility of production because the lawyers involved jointly agreed on the keyword search terms, the parties should save time and money en route to a resolution of the matters in controversy.

Even in *Da Silva Moore*, once the parties had agreed on a predictive coding, or machine-learning, approach to electronic searches, it is not too much of an extension of the keyword-search cooperation case law to ask the parties to try to reach agreement on the lessons being taught to the machine to retrieve responsive documents to avoid later disputes on the defensibility of the production.<sup>52</sup>

But does cooperation extend to capitulating to the demand of an opposing party to use a computer-search methodology that is allegedly superior to the one a party wishes to use because it is, say, more likely to find the “smoking gun” or the “hot” documents? That is the question lurking in the background in *Kleen Products*.

*Kleen Products* is an antitrust claim against several defendants. Plaintiffs accuse the defendants of price fixing. The district court judge referred to Judge Nolan the conduct of an evidentiary hearing to resolve a dispute over how defendants would search their electronic documents to respond to plaintiffs' requests for production.

At a February 21, 2012 hearing, an expert on the defendants' side defended the use of traditional Boolean keyword search approach on a software platform that allowed for additional analysis of documents by, for example, grouping documents by the words used in the documents and the frequency of their use.<sup>53</sup> An expert

---

<sup>51</sup> From the court docket, the parties' squabbles over the predictive coding protocol and the development of the seed set have been time consuming for the court and the parties. Many judges are not going to take the time to pore over individual documents to decide whether they should be included or excluded in a seed set. Among other results, the cost of achieving an acceptable seed set and the amount of court involvement in this matter will be scrutinized closely by the bench and the bar. There is another decision that will be scrutinized. In *Global Aerospace Inc. et al. v. Landow Aviation, L.P.*, Case No. CL 61040 (Cir. Ct. Loudin Co., Va. April 23, 2012), the circuit court allowed defendants to proceed “with the use of predictive coding for purposes of the processing and production of electronically stored information, with processing to be completed within 60 days and production to follow as soon as practicable and in no more than 60 days. This is without prejudice to a receiving party raising with the Court an issue as to completeness of the contents of the production or the ongoing use of predictive coding.”

<sup>52</sup> As discussed in the analysis of the introductory hypothetical below, however, this level of cooperation needs to be tempered by concerns over encroachment into lawyer work product. Courts also have to be on guard against converting cooperation into endless battles between dueling experts over the size and content of the seed set documents and relevance determinations being made by the producing party to populate the seed set. At some point, the producing party has to be able to move forward with a production and defend it if that ever becomes necessary once the production is made.

<sup>53</sup> This expert testified there was not much of a distinction between this approach and predictive coding: “Predictive coding is the ability to look at a few documents to make decisions, and to come up with a methodology, a technology that will take those decisions and apply them to a different body of documents. And that's what keywords do. When you look at documents and you say these keywords are choosing documents that I believe help me segregate my documents and find ones that are likely to be relevant and I apply those keywords which look at the full text of the documents I'm applying them against and segregate those

on the plaintiff's side faulted this search technique because it requires the user "to anticipate in advance of reviewing the documents the exact words that might be found in documents."<sup>54</sup> Relying in part on the results of the TREC Legal Track, this expert distinguished a "statistical ranked retrieval" or review<sup>55</sup> from a Boolean keyword review,<sup>56</sup> arguing the former was superior to the latter even if document clustering tools were being utilized.<sup>57</sup> The hearing was not concluded and was continued until March 28, 2012.

At the March 28, 2012 hearing, instead of hearing more testimony, Judge Nolan explained that she knew a lot more than she did in February and wanted to start over. She explained that she was a "believer" in Principle 6 of the Sedona Best Practices Recommendations & Principles for Addressing Electronic Document Production, which states that responding parties "are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically-stored information."<sup>58</sup> In her words, "I think the people who are producing the records, producing the documents are in a better position to know, since they have to do the work, spend the money, spend the time, they know their people, they know their material, so as a basic premise, I think that's a pretty fair premise here."<sup>59</sup>

She urged the experts on both sides to confer and reach a resolution that allowed defendants to proceed as they wished in a way that plaintiffs were comfortable that the search was statistically valid.<sup>60</sup>

In a subsequent hearing, plaintiffs' counsel advised the court that they had developed a protocol between predictive coding and a Boolean search that involved about 2,400 documents per defendant on which plaintiffs proposed to do random sampling to validate the quality of the search. Defendants objected to the proposal.<sup>61</sup> The judge urged further dialogue among counsel. At the next hearing, some of the defendants offered to produce a dictionary or index listing of all words used in the documents produced (and the other defendants

---

documents, it's making a decision." He added that his client was using "other types of content analytics in developing and testing their search terms" including a feature that uses "software technology to analyze the words in individual documents and create relationships between those words and the way those words appear on a frequency basis between multiple documents and in proximity to each other -- not proximity, but in co-appearance, if you will, in one document versus another. And again uses that to group documents, define similar documents, and make better decisions about how to segregate one set of documents from another. So from that aspect what the defendants have used is content based analytics, and they've used it in a predictive coding manner." Transcript of February 21, 2012 Proceedings, at 122-24, Docket Entry 304, March 1, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>54</sup> Transcript of February 21, 2012 Proceedings, at 234, Docket Entry 304, March 1, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>55</sup> Plaintiffs referred to their approach formally as a "Content Based Advanced Analytics" approach.

<sup>56</sup> "A statistical ranked retrieval system uses the statistical properties of language and can pay differing attention to different words. So again, the statistical ranked retrieval system is different from Boolean in that sense. Boolean, either the word is used or not in certain logical combinations. A statistical ranked retrieval system computes a numeric weight for each word and determines how much attention to pay to the word." Transcript of February 21, 2012 Proceedings, at 239, Docket Entry 304, March 1, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>57</sup> After describing what he believed were faults in one defendant's search protocol, he opined: "I conclude first that plaintiffs' proposal is likely to find a greater proportion of responsive documents with less human effort. And second, I conclude that plaintiffs' proposal will provide a statistically valid estimate of the recall of the information retrieval process." Transcript of February 21, 2012 Proceedings, at 243-257, 258, Docket Entry 304, March 1, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>58</sup> One might ask if Principle 6 conflicts with the Cooperation Proclamation. But as noted earlier there is a difference between choosing a search technology and then cooperating on its implementation and using an acceptable search methodology but then being told you must use a different one.

<sup>59</sup> Transcript of March 28, 2012 Proceedings, at 297-98, Docket Entry 319-1, May 17, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>60</sup> Transcript of March 28, 2012 Proceedings, at 297-98, Docket Entry 319-1, Exhibit A, May 17, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>61</sup> Transcript of April 2, 2012 Proceedings, Docket Entry 319-1, Exhibit B, May 17, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

agreed to see if their review systems could do the same). The list might show the number of “hits” for the word but if it did not, plaintiffs could follow up with defendants to run searches on a reasonable number of additional terms to determine the number of hits produced and then to follow up in an appropriate way depending upon the number of hits.<sup>62</sup> In their Joint Status Report No. 3 dated May 17, 2012, the parties stated that they were continuing to meet and confer in a good faith effort to resolve their discovery differences.<sup>63</sup> By stipulation dated August 21, 2012,<sup>64</sup> plaintiffs withdrew their request that defendants apply “Content Based Advanced Analytics” (CBAA) and agreed that they would not claim that defendants had a duty to use a search process other than the one defendants were already using with respect to the “First Request Corpus.”<sup>65</sup>

Reading the transcripts in this matter teaches an important lesson. Experts can defend any of a number of search technologies. In the absence of an agreement by litigants to use predictive coding or some other form of computer-assisted review, and without any evidence of a faulty production, courts are going to be very, very cautious before they even consider a request to dictate upfront what approach a producing party should take in making a production.<sup>66</sup>

This is not to say that lawyers can relax when they hear words like “Boolean connectors” or “statistical rank retrieval” or “Content Based Advanced Analytics.” To the contrary, a lawyer has a professional obligation to remain competent: “technology” amendments to the Model Rules of Professional Conduct have now been adopted.

## **AUGUST 2012 CHANGES TO THE MODEL RULES RELATED TO TECHNOLOGY**

The ABA Commission on Ethics 20/20 was formed to consider changes to the Model Rules of Professional Conduct with an eye in part on the intersection of lawyers’ conduct and advances in technology.<sup>67</sup> In August 2012, the ABA House of Delegates approved the Commission’s resolution on “technology and confidentiality” amending the Model Rules.<sup>68</sup>

One of the amendments relates specifically to a lawyer’s duty to provide competent representation. Comment [6] to Rule 1.1 addressing competence now includes the following highlighted clause: “[6] To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with relevant technology*, engage in continuing study and education and

---

<sup>62</sup> Transcript of April 19, 2012 Proceedings, Docket Entry 319-1, Exhibit C, May 17, 2012, Case No. 1:10-cv-05711 (N.D. Ill.)

<sup>63</sup> Docket Entry 319, May 17, 2012, Case No. 1:10-cv-05711 (N.D. Ill.).

<sup>64</sup> Docket Entry 385, August 21, 2012, Case No. 1:10-cv-05711 (N.D. Ill.)

<sup>65</sup> Plaintiffs further agreed that as to any ESI beyond the First Request Corpus, they would not contend that defendants should be required to use CBAA or any other technology that was the subject of the prior evidentiary hearings as to any requests for production served before October 1, 2013. For requests served after that date, the parties agreed to meet and confer regarding search methodologies and to file any motions thereafter as might be appropriate if they are unable to resolve any disputes regarding search methodologies.

<sup>66</sup> One way to view Judge Carter’s opinion in *Da Silva Moore* is that he was allowing MSL to choose its production methodology even if plaintiff objected to the protocol being used by MSL.

<sup>67</sup> See [http://www.americanbar.org/groups/professional\\_responsibility/aba\\_commission\\_on\\_ethics\\_20\\_20.html](http://www.americanbar.org/groups/professional_responsibility/aba_commission_on_ethics_20_20.html).

<sup>68</sup> For a discussion of the changes, see Barkett, *ABA to Tackle Technology Issues in Model Rules at August Meeting*, (<http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202560335059&thepage=3&slreturn=1>). Law Technology News (June 25, 2012).

comply with all continuing legal education requirements to which the lawyer is subject.” This provision will require lawyers to better understand any advances in technology that genuinely relate to competent performance of the lawyer’s duties to a client.

Search technology is also related to protecting client-confidential information from inadvertent disclosure. A change to Model Rule 1.6, addressing confidentiality of information, addresses this topic. Model Rule 1.6 now contains new subparagraph (c): “(c) A lawyer shall make reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client.”<sup>69</sup>

Another amendment is potentially significant where a client selects a vendor of computer-assisted review services and exercises control over the vendor’s work.

New Comment [4] to Model Rule 5.3 addresses the client’s (as opposed to the lawyer’s) selection of the nonlawyer service provider. Invoking Model Rule 1.2,<sup>70</sup> it explains that where the client picks the third-party service provider, “the lawyer ordinarily should agree with the client concerning the allocation of responsibility for monitoring as between the client and the lawyer.”<sup>71</sup> In a report accompanying these changes to Model Rule 5.3, the Commission wrote that, “In such situations, the lawyer ordinarily should consult with the client to determine how the outsourcing arrangement should be structured and who will be responsible for monitoring the performance of the nonlawyer services.”

The use of the word “monitoring” is new to the Model Rules. The Commission explained:

*The Commission concluded that [monitoring] was needed because, when a nonlawyer outside the firm is performing services in connection with a matter, it may not be possible to “directly supervise” the nonlawyer. The word “monitoring” makes clear that there is nevertheless a need to remain aware of how nonlawyer services are being perform[ed]. The Comment explains that, when the client directs the lawyer to use a particular nonlawyer, the lawyer and client should ordinarily agree who will have this “monitoring” responsibility. In contrast, if the client has not directed the selection of the nonlawyer, the lawyer or law firm would have the “monitoring” responsibility.*

---

<sup>69</sup> As noted earlier, Fed. R. Evid. 502(d) gives lawyers the ability to obtain a court order to protect a privileged or work-product protected document from waiver if it is inadvertently produced. One would think that in litigation a request to enter a 502(d) order is embraced by the words “reasonable efforts” in proposed Rule 1.6(c). In any event, from a malpractice protection standpoint alone, lawyers are ill-advised if they fail to seek a 502(d) order.

<sup>70</sup> Model Rule 1.2(a) is discussed below, but it provides that a lawyer “shall abide by a client’s decisions concerning the objectives of representation” and, as required by Model Rule 1.4, “shall consult with the client as to the means by which they are to be pursued.” A lawyer can limit the scope of the representation “if the limitation is reasonable under the circumstances and the client gives informed consent.” Model Rule 1.2(c). Under Model Rule 1.2(d), “A lawyer shall not counsel a client to engage, or assist a client, in conduct that the lawyer knows is criminal or fraudulent, but a lawyer may discuss the legal consequences of any proposed course of conduct with a client and may counsel or assist a client to make a good faith effort to determine the validity, scope, meaning or application of the law.”

<sup>71</sup> Proposed Comment [4] also states when a client and the lawyer allocate responsibility for monitoring the nonlawyer vendor, “lawyers and parties may have additional obligations that are a matter of law beyond the scope of these Rules.” In a separate report, the Commission explained this text “is intended to remind lawyers that they may have duties to a tribunal that are not necessarily satisfied through compliance with the Rules of Professional Conduct. For example, if a client instructs the lawyer to hire a particular electronic discovery vendor, the lawyer cannot cede all monitoring responsibility to the client, given that the lawyer may have to make certain representations to a tribunal regarding the vendor’s work.”

It seems unlikely that a lawyer would use the services of a predictive coding vendor and not be intimately involved in monitoring the work and work product of the vendor, especially in light of obligations placed on lawyers to properly manage discovery in litigation.<sup>72</sup>

Now that these changes have been adopted by the ABA House of Delegates, they will slowly be incorporated into State rules of professional conduct.

What is the likely consequence of these amendments? Within the ambit of reasonableness, lawyers will not be able to ignore advancements in search technology and must put themselves in a position to have an understanding of what technology works best in what situations and with what limitations, either on their own or through consultants.

## THE ETHICS OF COMPUTER-ASSISTED REVIEW: ANALYSIS OF THE INTRODUCTORY HYPOTHETICAL

Let's now turn to the current professional rules of conduct under this hypothetical. Without minimizing the importance to lawyers of staying abreast of technology, it is important to keep this analysis in perspective: the number of cases with sufficient amounts in controversy or volume of e-discovery to warrant the types of battles playing out in *Da Silva Moore* and *Kleen Products* are not large.<sup>73</sup>

---

<sup>72</sup> See discussion below on *Qualcomm Inc. v. Broadcom Corp.*, 2008 WL 66932 (S.D. Calif. Jan. 7, 2008) and *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 218 (S.D.N.Y. 2003).

<sup>73</sup> The Administrative Office of the United States Courts publishes annual statistics on court filings. For the one-year period ending March 31, 2011, there were 294,336 civil actions commenced in the federal courts. *U.S. District Courts--Civil Cases Commenced, by Basis of Jurisdiction and Nature of Suit, During the 12-Month Periods Ending March 31, 2010 and 2011*. <http://www.uscourts.gov/Viewer.aspx?doc=/uscourts/Statistics/FederalJudicialCaseloadStatistics/2011/tables/C02Mar11.pdf>. Of these, the United States was a plaintiff in 9,357 actions, or 3.2% of the cases, and it was a defendant in 35,833 actions, or 12.2% of the cases. Private actions represented 84.6% of the actions. By the nature of the suit according to forms completed by litigants at the time of filing, prisoner petition, social security, cable satellite TV, forfeiture and penalty, immigration, deportation, bankruptcy litigation, environmental, constitutional challenges to state statutes, "other statutory actions," tax suits, and Freedom of Information Action matters represented 30.3% of the filings. Civil rights (primarily employment) and labor law actions constituted 18.8% of the filings. Contract (primarily insurance and "other" but including franchise, marine, Miller Act, negotiable instruments, recovery of overpayments and enforcement of judgments, defaulted student loans, veterans' overpayments), and real property (condemnations, foreclosures, landlord tenant, torts to land, and "other") actions represented 14% of the filings. Tort actions (primarily asbestos and "other" but including airplane, marine, motor vehicle, assault, libel, slander, Federal Employers Liability Act, medical malpractice, and fraud including truth-in-lending) represented 29.5% of the cases commenced. Consumer credit actions represented 3.0% of the filings. One would expect that except for class actions, or claims affecting a large number of plaintiffs outside of a class action, these matters will not involve sophisticated electronic document search strategies. Intellectual property cases represent 3.3% of the filings, but almost two-thirds of these cases are trademark or copyright cases, which typically would not be expected to be large e-discovery matters. Antitrust, banking, RICO, and securities matters totaled 3,518 actions, or 1.2% of the filings. *Id.* *Da Silva Moore* is an employment action where class certification is being sought. *Kleen Products* is an antitrust case. In patent cases, the Federal Circuit's E-Discovery Committee has proposed an E-Discovery Model Order that focuses on email custodians and search terms and limits the number of both. [http://www.cafc.uscourts.gov/images/stories/announcements/Ediscovery\\_Model\\_Order.pdf](http://www.cafc.uscourts.gov/images/stories/announcements/Ediscovery_Model_Order.pdf). The Eastern District of Texas, where a large number of patent infringement lawsuits are filed, has edited this Model Order and made it an Appendix P to its Local Rules but still limits the number of email custodians and search terms. <http://www.txed.uscourts.gov/page1.shtml?location=rules>. See also *Walking the Plank, Looking Over Your Shoulder, Fearing Sharks are in the Water: E-Discovery in Federal Litigation?* (Duke 2010 Conference, Civil Rules Advisory Committee, May 11, 2010) (<http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Duke%20Materials/Library/John%20Barkett,%20Walking%20the%20Plank.pdf>), at 5-8 and accompanying footnotes (discussing the results of various surveys of lawyers and describing three types of cases requiring strong case management: the symmetric case (data-rich party against data-rich party); the asymmetric case (a data-poor party against a data-rich party trying to take advantage of the disparity); and the disproportionate case (where irrespective of data, the amount or issues in controversy are too small in relation to e-discovery costs).

## Competence

Model Rule 1.1 covers competence. It provides:

*A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation.*

There is no doubt that lawyers engaged in meet-and-confer sessions with opposing counsel must be prepared to deal with search protocol issues. When judges are demanding cooperation, competent lawyers must know this. Similarly, when search protocol disputes appear to be a focal point of judicial anger over the lack of cooperation competent lawyers must take notice.

Similarly, competence also means knowing how to prepare for the meet-and-confer session. Whether the issue is a search protocol, or databases, storage practices, recycling of backup tapes, or key players, a lawyer has to know what positions the lawyer will take and must have a sense of how the judge will react to those positions should the parties fail to reach agreement on how to proceed.

Search technology and capability is evolving. The Legal Track at TREC demonstrates this fact. Competent lawyers, or persons who work for or are retained by competent lawyers, must stay abreast of the actual performance of computer-assisted review and be in a position to weigh the costs and benefits of different approaches to satisfying a request for production or searching for documents in support of a claim or a defense.<sup>74</sup> Lawyers handling e-discovery must, at a minimum, have an awareness of the general issues involved in evaluating different search methodologies, and know where to look or whom to turn to determine if computer-assisted review should be employed for a particular e-discovery production.

Is a lawyer qualified to advance the use of search protocols that go beyond search terms and involve linguists and computer scientists designing searches that improve both the rate of recall (finding relevant information) and precision (finding *only* relevant information)? Competent lawyers must know enough to advance support for the approach or to question it, depending upon one's position. What if opposing counsel appears with an expert to support the search approach? Can you ask to cross examine the witness? Must you have a witness of your own? Can you argue against a search approach without an expert in light of cases like *O'Keefe* and *Victor Stanley* despite Judge Peck's statement in *Da Silva Moore* that "Rule 702 and Daubert are simply not applicable to how documents are searched for and found in discovery"? The answers to these questions will likely vary from judge to judge.

Competence is also implicated in the development of an ESI protocol. In the ESI protocol that the court approved in *Da Silva Moore*, MSL agreed to provide to plaintiffs all of the documents reviewed "as a function of the seed set," including all nonprivileged documents marked as nonresponsive. Nonresponsive documents that were pulled in the sample of the document collection used to form the seed set could contain evidence that is embarrassing to an employee (e.g., when no one knows an employee is having marital issues, an email from that employee making an appointment to see a marriage counselor that happens to contain a key word used to locate documents for the seed set), harmful to the party (evidence of an environmental violation), or

---

<sup>74</sup> So many e-discovery vendors have sprung up in the past few years that lawyers who are not cautious will regret the lack of caution. If predictions about success in cutting e-discovery costs sound too good to be true, they likely are not true.

information which the requesting party might recognize would be useful in an unrelated pending or potential cause of action. The same privacy concerns that animate the protections contained in protocols for forensic examinations of personal computers<sup>75</sup> must be on the minds of lawyers before they agree to open-ended review of documents coded as nonresponsive.<sup>76</sup>

Where a lawyer is handling discovery, e-discovery and evolving search technology are redefining “competence” for that lawyer.<sup>77</sup>

### *Diligence*

Model Rule 1.3 provides: “A lawyer shall act with reasonable diligence and promptness in representing a client.”

Comment [1] to Rule 1.3 discusses the dedication and “zeal” with which a lawyer is to represent a client:

*A lawyer should pursue a matter on behalf of a client despite opposition, obstruction or personal inconvenience to the lawyer, and take whatever lawful and ethical measures are required to vindicate a client’s cause or endeavor. A lawyer must also act with commitment and dedication to the interests of the client and with zeal in advocacy upon the client’s behalf.*

Is the emphasis on cooperation inconsistent with the lawyer’s duty to diligently represent the lawyer’s client? Principle 1.02 of the Seventh Circuit’s Pilot Project Principles, quoted above and again here, specifically says it is not:

*An attorney’s zealous representation of a client is not compromised by conducting discovery in a cooperative manner. The failure of counsel or the parties to litigation to cooperate in facilitating and reasonably limiting discovery requests and responses raises litigation costs and contributes to the risk of sanctions.<sup>78</sup>*

---

<sup>75</sup> See *Ameriwood Indus. Inc. v. Liberman*, 2006 U.S. Dist. LEXIS 93380 (E.D. Mo. Dec. 27, 2006) (allowing forensic examination of personal computers but allowing producing party to withdraw documents for privacy or privilege reasons); See, generally, Barkett, *E-Discovery: Twenty Questions and Answers* (Chicago: First Chair Press, 2008), at 13-20 (discussing discovery on e-discovery).

<sup>76</sup> For an extensive ESI/“advanced analytics,” predictive coding protocol in a less-publicized matter, see Case Management Order: Protocol Relating to the Production of Electronically Stored Information,” *In Re Actos (Pioglitazone) Products Liability Litigation*, MDL No. 6:11-md-2299 (W.D. La. July 27, 2012) (the order can be found at <http://pdfserver.amlaw.com/legaltechnology/11-md-2299.pdf>).

<sup>77</sup> As noted earlier, Fed. R. Civ. P. 26(g)(1) addresses initial disclosures under Fed. R. Civ. P. 26(a) and states: “Every disclosure under Rule 26(a)(1) or (a)(3) and every discovery request, response, or objection must be signed by at least one attorney of record in the attorney’s own name.... By signing, an attorney or party certifies that to the best of the person’s knowledge, information, and belief formed after a reasonable inquiry: (A) with respect to a disclosure, it is complete and correct as of the time it is made;... Under Rule 26(g)(c), if “without substantial justification,” the certification is made in violation of Rule 26(g)(1)(A), the district court upon motion or on its own must impose an appropriate sanction on the signer, the party on whose behalf the signer was acting, or both. The sanction may include an order to pay the reasonable expenses, including attorneys’ fees, caused by the violation. Will the day come when the search approach taken might implicate these rules where recall and precision could be improved by using a different search approach? We are not there yet, and cost cannot be ignored in evaluating the answer to this question.

<sup>78</sup> Comment [1] to Model Rule 1.3, adds this qualification to the lawyer’s duty of zealous advocacy: “A lawyer is not bound, however, to press for every advantage that might be realized for a client. For example, a lawyer may have authority to exercise professional discretion in determining the means by which a matter should be pursued. See Rule 1.2.” Model Rule 1.2 is discussed further below.

Lawyers must properly balance client needs with calls for cooperation. A lawyer is ill-advised to refuse to discuss search terms or search protocols with opposing counsel who wants to attempt to reach agreement on these issues. Where a lack of cooperation puts a client at risk of sanctions or higher litigation costs, lawyers have to take notice.<sup>79</sup>

On the other hand, where a judge makes demands that are unreasonable, lawyers have to protect their clients' interests. Rules 1.1 and 1.3 require lawyers to know when to cooperate and when to stand their ground on an e-discovery question.

The ESI protocol approved in *Da Silva Moore* gave the requesting party access to the responsiveness judgments and issue tags used by MSL.<sup>80</sup> Such transparency highlights the tension between cooperation and advocacy, or to use another phrase, between case management and client confidences. Lawyers will argue that the decisions made about responsiveness generally and in relation to issues in controversy are opinion work product.<sup>81</sup> Judge Peck appeared to realize that he was trading into work product territory when he wrote: "While not all experienced ESI counsel believe it necessary to be as transparent as MSL was willing to be, such transparency allows the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called 'black box' of the technology. This Court highly recommends that counsel in future cases be willing to at least discuss, if not agree to, such transparency in the computer-assisted review process." 2012 U.S. Dist. LEXIS 23350. Judge Peck's recommendation notwithstanding, in the last

---

<sup>79</sup> See *Covad Communications Co. v. Revonet, Inc.*, 2010 U.S. Dist. LEXIS 31165, \*50 (D.D.C. Mar. 31, 2010) (where the magistrate judge awarded the fees incurred by defense counsel to try to reach a discovery accord because plaintiff's counsel failed to engage in meaningful discussions to resolve the discovery dispute).

<sup>80</sup> The ESI protocol provided: "MSL will provide Plaintiffs' counsel with all of the non-privileged documents and will provide, to the extent applicable, the issue tag(s) coded for each document . . . . If necessary, counsel will meet and confer to attempt to resolve any disagreements regarding the coding applied to the documents in the seed set." 2012 U.S. Dist. LEXIS 23350, \*36-37.

<sup>81</sup> Where a compilation or selection process reflects the mental impressions or legal strategy of counsel, it would be opinion work product. *Lockheed Martin Corp. v. L-3 Communications Corp.*, 2007 WL 2209250, \*10 (M.D. Fla. July 29, 2007) (documents "containing instructions about how to conduct the search and what specifically to search for are opinion work product"); *Burroughs Wellcome Co. v. Barr Laboratories, Inc.*, 143 F.R.D. 611, 624 (E.D.N.C. 1992) (where defendant sought the printed results of database searches conducted at the behest of plaintiff's counsel, work product protection was afforded the compilation: "The court agrees with plaintiff that the compilation of search results reflects the legal strategy of counsel.") See generally *Sporck v. Peil*, 759 F.2d 312, 315 (3d Cir. 1985) (where plaintiff sought certain nonprivileged documents which defense counsel had selected out a large document production to prepare a defense witness for a deposition, because "identification of the documents as a group will reveal defense counsel's selection process, and thus his mental impressions," identification of the documents as a group "must be prevented to protect defense counsel's work product.") Where disclosure of a selection and compilation of documents would not reveal counsel's thought processes, "ordinary" or fact work product only may exist. See *In Re San Juan Dupont Plaza Hotel Fire Litigation*, 859 F.2d 1007, 1019 (1st Cir. 1988) (distinguishing between a case management order and a discovery order in a matter that will involve 2,000 depositions, upholding a requirement that the parties provide a list of documents they would use to interrogate a deponent five days before the deposition because they represented ordinary, not opinion, work product. The court of appeals found no abuse of discretion in the district court's case management order that explained that "the parties had manifested a substantial need for deposition exhibit lists arising out of the time and delay inherent in reviewing and raising objections to deposition exhibits during, rather than before, each deposition." The court of appeals held that the order "had a systemically beneficial component: not only would litigants' burdens be eased, but the court's ability to administer, process, and respond adequately to the case's ebb and flow would be materially enhanced." It also cited the Manual for Complex Litigation which endorsed the prior identification of deposition exhibits as a way to facilitate discovery); *Disability Rights Council of Greater Washington v. Washington Metropolitan Transit Auth.*, 242 F.R.D. 139, (2007) (where plaintiffs subpoenaed from a contractor customer complaints lodged by disabled riders during the term of the contractor's contract with WMATA and searched through over forty boxes of documents to collect complaints consistent with plaintiffs' theory of the case, WMATA's request for production of the complaints was rejected. The selection process represented fact, not opinion, work product because there were thousands of documents involved and "it would be difficult to conceive that Plaintiffs' trial strategy could be gleaned solely by virtue of Plaintiffs' disclosure of the documents selected." However, WMATA did not show either substantial need or undue hardship required to obtain fact work product under Fed. R. Civ. P. 26(b)(3)). See also *Gould v. Mitsui Mining & Smelting Co., Ltd.*, 825 F.2d 676, 680 (2d Cir. 1987) (calling the exception in *Sporck* "narrow" and remanding a decision to quash a subpoena served on counsel to a party to require the district court to determine whether there was "a real, rather than speculative, concern that the thought processes" of counsel would be exposed).

part of the hypothetical, the lawyer facing demands for such transparency may be unable to acquiesce in order to protect a client's interests and the lawyer's work product.

In e-discovery, there is now an additional premium placed on preparation. We are in the midst of evolving technologies that may be embraced by some lawyers sooner than other lawyers. Lawyers must be prepared when opposing counsel suggests using search protocols that do not involve human reviewers.

### *Expediting Litigation*

Model Rule 3.2 is rarely invoked but, where cooperation with respect to e-discovery is an issue, it also may be applicable to a lawyer's conduct. It provides in full:

*A lawyer shall make reasonable efforts to expedite litigation consistent with the interests of the client.*

The comment to this rule explains: "Dilatory practices bring the administration of justice into disrepute."<sup>82</sup> A lawyer who is not cooperative in e-discovery may not be engaging in a dilatory practice, but if the effect of the lawyer's conduct violates the "reasonable efforts" clause in Model Rule 3.2, the lawyer should not be surprised if a judge invokes this rule to question the lawyer's conduct.

### *Communication*

The first two parts to the hypothetical illustrate why communication with the client is essential if e-discovery is going to be managed. The Rules of Professional Conduct emphasize client communication. Model Rule 1.2 provides in pertinent part that "a lawyer shall abide by a client's decisions concerning the objectives of representation and, as required by Rule 1.4, shall consult with the client as to the means by which they are to be pursued. A lawyer may take such action on behalf of the client as is impliedly authorized to carry out the representation." Model Rule 1.4(a) provides in part that a lawyer shall "(2) reasonably consult with the client about the means by which the client's objectives are to be accomplished"; "(3) keep the client reasonably informed about the status of the matter"; and "(5) consult with the client about any relevant limitation on the lawyer's conduct when the lawyer knows that the client expects assistance not permitted by the Rules of Professional Conduct or other law." Model Rule 1.4(b) provides that "a lawyer shall explain a matter to the extent reasonably necessary to permit the client to make informed decisions regarding the representation." Model Rule 2.1 provides: "In representing a client, a lawyer shall exercise independent professional judgment and render candid advice."

In the first part of the hypothetical, the lawyer must inform the general counsel of the risks the client faces because of the potential unreasonableness of the position being taken on discussion with opposing counsel of search terms. Independent legal judgment, candid advice, and proper explanation of the risks and benefits of an e-discovery position, represent professional conduct.

In the second part of the hypothetical, the lawyer might have taken opposing counsel's request under advisement. The lawyer might have asked to learn more about the search technology and whether it has been

---

<sup>82</sup> The comment concludes: "The question is whether a competent lawyer acting in good faith would regard the course of action as having some substantial purpose other than delay. Realizing financial or other benefit from otherwise improper delay in litigation is not a legitimate interest of the client."

actually employed in other litigation with or without judicial approval and with what results. The lawyer might have confessed to a lack of knowledge of the TREC Legal Track and asked for a chance to read *Da Silva Moore* and then a resumption of the meet-and-confer session. The lawyer might have also asked to speak with the vendor proposed for use by opposing counsel, perhaps with a vendor of the lawyer present as well, to discuss the vendor's track record and the protocol proposed by the vendor and whether the vendor participated in the TREC Legal Track. The lawyer then should discuss with the client what opposing counsel has proposed and evaluate possible responses. Irrespective of the communication obligations under the Model Rules, there is no shame in reserving judgment, learning, communicating with the client, developing a strategy, and then reconvening the meet-and-confer session.

### *Supervision*

Model Rule 5.1 requires in part that a partner in a law firm and lawyers engaged in management of the law firm to make reasonable efforts to ensure that the firm "has in effect measures giving reasonable assurance that all lawyers in the firm conform to the Rules of Professional Conduct." Competence is the first rule of professional conduct, and embraces sufficient knowledge of search capability and technology suitable to meet the e-discovery needs of each matter handled by a lawyer.

*(a) A partner in a law firm, and a lawyer who individually or together with other lawyers possesses comparable managerial authority in a law firm, shall make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that all lawyers in the firm conform to the Rules of Professional Conduct.*

*(b) A lawyer having direct supervisory authority over another lawyer shall make reasonable efforts to ensure that the other lawyer conforms to the Rules of Professional Conduct.*

*(c) A lawyer shall be responsible for another lawyer's violation of the Rules of Professional Conduct if:*

*(1) the lawyer orders or, with knowledge of the specific conduct, ratifies the conduct involved;  
or*

*(2) the lawyer is a partner or has comparable managerial authority in the law firm in which the other lawyer practices, or has direct supervisory authority over the other lawyer, and knows of the conduct at a time when its consequences can be avoided or mitigated but fails to take reasonable remedial action.*

Model Rule 5.3 requires lawyers to be able to supervise nonlawyers, and that would include e-discovery vendors. It provides:

*With respect to a nonlawyer employed or retained by or associated with a lawyer:*

*(a) a partner, and a lawyer who individually or together with other lawyers, possesses comparable managerial authority in a law firm shall make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that the person's conduct is compatible with the professional obligations of the lawyer;*

*(b) a lawyer having direct supervisory authority over the nonlawyer shall make reasonable efforts to ensure that the person's conduct is compatible with the professional obligations of the lawyer; and*

*(c) a lawyer shall be responsible for conduct of such a person that would be a violation of the Rules of Professional Conduct if engaged in by a lawyer if:*

*(1) the lawyer orders or, with the knowledge of the specific conduct, ratifies the conduct involved; or*

*(2) the lawyer is a partner or has comparable managerial authority in the law firm in which the person is employed, or has direct supervisory authority over the person, and knows of the conduct at a time when its consequences can be avoided or mitigated but fails to take reasonable remedial action.*

The buck has to stop with the supervising lawyer when a third-party e-discovery vendor is involved and is being managed by the lawyer. *Cf. Mays v. Neal*, 938 S.W.2d 830, 835 (Ark. 1997) (invoking Mo. RPC 5.3(b) and saying: “while a lawyer may delegate certain tasks to his assistants, he or she, as supervising attorney, has the ultimate responsibility for compliance by the nonlawyer with the applicable provisions of the Model Rules”); *In Re Flack*, 33 P.3d 1281, 1287 (Kan. 2001) (a lawyer who contracted with a probate services company violated Kansas RPC 5.3 “when he failed to properly supervise the client service representatives and when he failed to assure that the actions of the client service representatives were compatible with his professional obligations as a lawyer.”)

A party is responsible for its e-discovery vendor's errors. See *In Re Seroquel Products Liability Litigation*, 2007 U.S. Dist. LEXIS 61287, \*49-50 (M.D. Fla. Aug. 21, 2007) (a party is responsible for the errors of its vendors citing Sedona Principle 6: “Ultimate responsibility for ensuring the preservation, collection, processing, and production of electronically stored information rests with the party and its counsel, not with the nonparty consultant or vendor.”) If a lawyer assumes the management of a third-party vendor, it will behoove the lawyer to know what else the Sedona Principles say about vendors, since these principles are often cited by courts in e-discovery matters:

*Due to the complexity of electronic discovery, many organizations rely on consultants to provide a variety of services, including discovery planning, data collection, specialized data processing, and forensic analysis. Such consultants can be of great assistance to parties and courts in providing technical expertise and experience with the collection, review, and production of electronically stored information. However, parties should carefully consider the experience and expertise of a potential consultant before his or her selection, as standards for experts and consultants in this field have not been fully developed. Vendors offer a variety of software and services to assist with the electronic discovery process and a party's evaluation of vendor software and services should include the defensibility of the process in the litigation context, the cost, and the experience of the vendor. Ultimate responsibility for ensuring the preservation, collection, processing, and production of electronically stored information rests with the party and its counsel, not with the nonparty consultant or vendor.*

*At all times, counsel, clients, and vendors must understand everyone's role in the discovery process. Thus, even if a vendor is retained to serve in a non-testifying capacity, everyone*

*should be aware of the potential need for testimony if forensic or other technical expertise is required to prepare electronically stored information for review or production. Additionally, care should be taken to ensure that the vendor does not assume the role of a legal advisor, and that all persons involved understand what communications are protected under the attorney-client privilege and what information may be protected as attorney work product.*<sup>83</sup>

In the last part of the hypothetical, counsel for the producing party is relying heavily on her vendor for support. That means hiring competent vendors and doing enough due diligence before and during the relationship to be comfortable that the vendor is serving the client's best interests.

Lawyers must be mindful of the breadth of their duty to be competent under Rule 1.1 as well as their obligations under Rule 5.3 before undertaking the responsibility to manage a third-party e-discovery vendor particularly in the realm of computer-assisted review.<sup>84</sup>

### *Duty of Candor Toward the Tribunal*

Model Rule 3.3(a)(1) and (3) are part of the duty of candor that every lawyer owes to a court. They provide:

*(a) A lawyer shall not knowingly:*

*(1) make a false statement of fact or law to a tribunal or fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer; ...*

*(3) offer evidence that the lawyer knows to be false. If a lawyer, the lawyer's client, or a witness called by the lawyer, has offered material evidence and the lawyer comes to know of its falsity, the lawyer shall take reasonable remedial measures, including, if necessary, disclosure to the tribunal. A lawyer may refuse to offer evidence, other than the testimony of a defendant in a criminal matter, that the lawyer reasonably believes is false.*

---

<sup>83</sup> [http://www.thesedonaconference.org/content/miscFiles/TSC\\_PRINCP\\_2nd\\_ed\\_607.pdf](http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf) (at 40).

<sup>84</sup> The dollars involved in dealing with e-discovery vendors can be large. To illustrate, Sullivan & Cromwell sued a vendor for alleged mistakes in handling e-discovery issues. In a complaint filed December 28, 2007, in the Southern District of New York, S&C accused Electronic Evidence Discovery, Inc. (EED) of "untimely and inaccurate" work. S&C asked the district court to declare that it did not have to pay a \$710,000 bill from EED. See <http://www.law.com/jsp/article.jsp?id=1199441137204>. EED responded by suing S&C in King County Superior Court in Washington on January 7, 2008, demanding payment of \$660,016.17 plus interest in the amount of \$58,592.07. The matter was quickly and confidentially settled according to a January 18, 2008 report in the New York Law Journal. Disputes can also arise where a vendor is terminated in the middle of litigation and raises cost demands before returning the client's electronically stored information. Vendors do make mistakes. *PSEG Power New York Inc. v. Alberici Constructors*, 2007 U.S. Dist. Lexis 66767 (N.D.N.Y. Sept. 7, 2007) (vendor separated 3000 emails from their attachments); *In Re Seroquel Products Liability Litigation*, 2007 U.S. Dist. LEXIS 61287, \*37-38 (M.D. Fla. Aug. 21, 2007) ("a significant portion of the production had blank pages; new load files were not searchable, in part because the date formats in the metadata were inconsistently loaded and email attachments not consistently associated or identified; authors were not identified as custodians for files; transposed metadata recipients/authors; and no page breaks were inserted in 3.75 million pages.") The e-discovery trade press also is following *J-M Mfg. Co., Inc. v. McDermott Will & Emery et al.*, Case No. BC 462 832 (Superior Ct., Los Angeles Co. Cal.) where plaintiff alleges in part that the law firm negligently supervised contract attorneys and vendors doing document review resulting in the production of privileged documents. Second Amended Complaint (June 25, 2012), at 13.

Does a party that used computer-assisted review to make a production without the fact of its use ever coming up in discussions with opposing counsel or the court have an ethical duty to disclose this fact to the court? The answer is “no.” However, the lawyer for that party will have to consider the lawyer’s obligations under Model Rule 1.2 to explain to the client the costs and risks associated with later discovery of the use of computer-assisted review that might, for tactical reasons, become an issue of contention and why it may be prudent to raise the planned use of computer-assisted review upfront with the court and opposing counsel.

In federal court, at least, if a Rule 16(b) conference is held,<sup>85</sup> it may result in questioning about e-discovery production protocols that might implicate the ethical duty of candor to the court.<sup>86</sup> Apart from the sanctions’ risk associated with misrepresentations to a court,<sup>87</sup> if a Rule 16(b) scheduling conference is held and a lawyer is considering advancing predictive coding, or any other form of computer-assisted review as an e-discovery protocol, the lawyer will have to consider the transparency singled out by Judge Peck as so important in his decision in *Da Silva Moore*, especially if, like in the hypothetical, the parties are not in agreement. The party proponent must not promise more than can be delivered, any more than the party opposing the use of predictive coding should engage in hyperbole or exaggeration over the discovery-failure risks posed by the use of the technology. And the lawyer advancing the use of computer-assisted review will need to consider how much the lawyer is able to disclose because of the lawyer’s duty of confidentiality.

### *Duty of Confidentiality*

Model Rule 1.6 requires lawyers to maintain the confidentiality of client information:

*(a) A lawyer shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is permitted by paragraph (b).*

Model Rule 1.6(b)(5) permits disclosure “to the extent the lawyer reasonably believes necessary”:

*(5) to establish a claim or a defense in a controversy between the lawyer and the client, to establish a defense in a criminal or civil claim against the lawyer based upon conduct in*

---

<sup>85</sup> Rule 16(b) provides for but does not require a scheduling conference.

<sup>86</sup> Fed. R. Civ. P. 16(b)(3) provides that a scheduling order may provide for “the disclosure of electronically stored information.”

<sup>87</sup> See, e.g., *The Pension Committee of the University of Montreal Pension Plan et al. v. Banc of America Securities, LLC et al.*, 685 F. Supp. 2d 456 (S.D.N.Y. 2010) where misrepresentations played a role in the district court’s decision to issue a “spoliation charge” to the jury. The opinion is discussed at length in Barkett, *Zubulake Revisited: Pension Committee and the Duty to Preserve* (ABA Section of Litigation News, February 26, 2010) ([http://www.abanet.org/litigation/litigationnews/trial\\_skills/pension-committee-zubulake-ediscovery.html](http://www.abanet.org/litigation/litigationnews/trial_skills/pension-committee-zubulake-ediscovery.html)). See also *Qualcomm Inc. v. Broadcom Corp.*, 2008 WL 66932 (S.D. Calif. Jan. 7, 2008) where a failure to produce relevant documents was disclosed after a jury trial and several lawyers, after a sanctions hearing, were referred to the California Bar for, in part, a breach of California’s equivalent to the Model Rules’ duty of candor. This sanction was later vacated to give the lawyers the chance to use privileged documents to defend positions taken before the district court, and remanded for a new hearing on sanctions, 2008 WL 638108,\*3 (S.D. Calif. March 5, 2008). No sanction was entered after the rehearing. *Qualcomm Inc. v. Broadcom Corp.* 2010 U.S. Dist. LEXIS 33889 (S.D. Cal. Apr. 2, 2010) (failure of outside counsel to meet with key players to explain document collection requirements; comprehend Qualcomm’s computer system; and supervise proper collection of electronically stored information did not amount to bad faith required by *Chambers v. NASCO, Inc.* 501 U.S. 32 (1991): “Sanctions only may be imposed under the court’s inherent authority upon a finding that the attorney acted in bad faith”). No lawyer would want to go through two years of professional hell, of course, until an outcome like this was reached. And the basis for this decision is of no consequence to lawyers in jurisdictions that do not require bad faith to impose sanctions for a violation of a court’s inherent authority. See *Walking the Plank, Looking Over Your Shoulder, Fearing Sharks Are in the Water: E-Discovery in Federal Litigation?* (Duke 2010 Conference, Civil Rules Advisory Committee, May 11, 2010) (<http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Duke%20Materials/Library/John%20Barkett,%20Walking%20the%20Plank.pdf>), at 29-33 (discussing the various culpability standards in the federal circuits).

*which the client was involved, or to respond to allegations in any proceeding concerning the lawyer's representation of the client.*<sup>88</sup>

Under the hypothetical, there are, at least, three ethical concerns here. First, how much information should the proponent of predictive coding provide before there is a risk of revealing client-confidential information? This is a variant on the obligation to communicate with the client, of course, but a lawyer, nonetheless needs to be thoughtful in what is said if the lawyer is advocating a predictive coding review.

Second, the duty of candor under Model Rule 3.3 applies “even if compliance requires disclosure of information otherwise protected by Rule 1.6.” Model Rule 3.3(c). In other words, the duty of candor to a tribunal trumps the duty of confidentiality. Lawyers need to be always mindful of this fact as they move between information learned from their clients and information provided to the court. This is particularly true if the client, not the lawyer, is managing the predictive coding review.

Third, if privileged or work-product protected information might be disclosed as part of computer-assisted review, entry of an order under Fed. R. Evid. 502(d) insulating the information from waiver in the event of disclosure, not just in the proceeding in which the order is entered but also in any other federal or state proceeding, must be a part of the tool kit that is used by the lawyer.

#### *Fairness to Opposing Party and Counsel*

Model Rule 3.4(a) provides:

*A lawyer shall not:*

*(a) unlawfully obstruct another party's access to evidence or unlawfully alter, destroy or conceal a document or other material having potential evidentiary value. A lawyer shall not counsel or assist another person to do any such act; ...*

It seems unlikely that this model rule would be invoked to argue that computer-assisted review *must* be used—at least not anytime soon. But will we approach the day when the argument will be made that the failure to use predictive coding or other forms of computer-assisted review, in fact, obstructs another party's access to evidence? That's a scary thought. But as data volume continues to grow and for the right type of case, it will become less and less of a fanciful thought.

#### *Reporting Professional Misconduct*

Rule 8.3(a) is a treacherous rule for all lawyers. It provides:

*(a) A lawyer who knows that another lawyer has committed a violation of the Rules of Professional Conduct that raises a substantial question as to that lawyer's honesty, trustworthiness or fitness as a lawyer in other respects, shall inform the appropriate professional authority.*

---

<sup>88</sup> Hence, in a sanctions dispute between the outside counsel and the in-house counsel as occurred in *Qualcomm Inc. v. Broadcom Corp.*, *supra*, under the Model Rules, outside counsel could have submitted attorney-client communications in an effort to defend themselves from an imposition of sanctions.

There are exceptions. For example, Model Rule 8.3(c) provides in pertinent part: “This Rule does not require disclosure of information otherwise protected by Rule 1.6...”

There is obviously an enormous amount of judgment involved in applying Model Rule 8.3. However, if the change to Model Rule 1.1 is made and adopted by the state in which the litigation is being handled, and the lawyer is not staying abreast of technology changes, should opposing counsel report the lawyer to the Bar? It seems unlikely that the opposing lawyer would do so today but in the coming years, “fitness” may take on new meaning as technology takes on more prominence in a lawyer’s practice.

#### *Potential Litigation Concern: Certification*

Fed. R. Civ. P. 26(g)(1) addresses initial disclosures under Fed. R. Civ. P. 26(a) and states:

*Every disclosure under Rule 26(a)(1) or (a)(3) and every discovery request, response, or objection must be signed by at least one attorney of record in the attorney’s own name.... By signing, an attorney or party certifies that to the best of the person’s knowledge, information, and belief formed after a reasonable inquiry:*

*(A) with respect to a disclosure, it is complete and correct as of the time it is made; and*

*(B) with respect to a discovery request, response, or objection, it is:*

*(i) consistent with these rules and warranted by existing law or by a nonfrivolous argument for extending, modifying, or reversing existing law, or for establishing new law;*

*(ii) not interposed for any improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation; and*

*(iii) neither unreasonable nor unduly burdensome or expensive, considering the needs of the case, prior discovery in the case, the amount in controversy, and the importance of the issues at stake in the action.*

Violations of Rule 26(g) are sanctionable. Under Rule 26(g)(3), if “without substantial justification,” the certification is made in violation of Rule 26(g), the district court “upon motion or on its own, *must impose* an appropriate sanction on the signer, the party on whose behalf the signer was acting, or both. The sanction may include an order to pay the reasonable expenses, including attorney’s fees, caused by the violation.” (Emphasis added.) *See Cache La Poudre Feeds, LLC v. Land O’Lakes, Inc. et al.*, 2007 U.S. Dist. LEXIS 15277, \*56-57, 80-81 (D. Colo. Mar. 2, 2007) (citing the Advisory Committee Note to Rule 26(g) on the obligation of counsel affirmatively to engage in pretrial discovery in a responsible manner consistent with the spirit and purposes of Rules 26-37, and the mandatory sanction language of Rule 26(g), the district court sanctioned the defendant, although because of the absence of prejudice, the sanction was \$5,000.00 plus the costs of a court reporter and the transcription of one deposition); *Wingnut Films Ltd. v. Katja Motion Pictures Corp. et al.*, 2007 U.S. Dist. Lexis 72953 (C.D. Cal. Sept. 18, 2007) (failure of counsel to make a reasonable investigation and effort to certify that client had provided all information and documents available to it resulted in a Rule 26(g) sanction).

Reference to Rule 26(g) may be putting the cart before the horse. But lawyers in federal court evaluating the use of computer-assisted review should be mindful of the obligations of Rule 26(g) as they plan their communications with opposing counsel and the court.

### *Potential Litigation Concern: Inherent Authority of a Court to Impose Sanctions*

Courts have inherent authority to sanction lawyers and clients. *Phoenix Four, Inc. v. Strategic Resources Corp. et al.*, 2006 WL 1409413 (S.D.N.Y. May 23, 2006). In this case, the district court found a law firm failed to locate and in a timely manner produce evidence stored in a server that the law firm's client (defendants) took with it when it moved from one building (leaving a defunct operation) to another. An electronic search was made of the computer system in defendants' new offices and no relevant electronically stored information was located. However, the servers that had been moved from the defunct operation were not searched. *Id.* at \*2. By coincidence, there was a server malfunction at the new offices and a technician was hired to solve it. The technician found a dormant drive containing relevant electronically stored information. Litigation counsel, Mound Cotton, was notified and promptly produced the electronic data representing 200-300 boxes of documents. *Id.* Plaintiff filed a motion for sanctions because of the late production. The district court agreed and imposed a sanction because the law firm had never undertaken a methodical survey of sources of electronically stored information. *Id.* at \*5-6. Both the defendants and the law firm were ordered to reimburse Phoenix equally for costs and fees associated with the filing of the motion for sanctions. The district court had to approve the amount once time records were produced. They were ordered to pay \$10,000 each for the re-depositions of three witnesses for the limited purpose of inquiring into issues raised by the documents recovered from the server. The district court also ordered that the defendants' shares could not be paid by their insurers. *Id.* at \*9.

Cases like this should emphasize the importance of avoiding mistakes that can jeopardize the value of computer-assisted review. As should be obvious from *Da Silva Moore*, if the search model is not properly constructed and managed, the search software will not produce optimal results. The custodians whose data will be reviewed and other data sources, both topics considered in *Da Silva Moore*, while always the subject of thoughtful consideration, become arguably more important to courts evaluating requests to conduct predictive coding review as a way to limit the costs of production.

### *Potential Litigation Concern: The Duty to Monitor Production*

*Zubulake v. UBS Warburg LLC*, 229 F.R.D. 422 (S.D.N.Y. 2004) (*Zubulake V*) sets forth an outside counsel's duty to monitor production by the client. I attach as an Appendix the language from *Zubulake V* on this subject, which bears re-reading since it may play an important role in gathering information necessary to draft a reasonable computer-assisted review protocol.

### *Potential Litigation Concern: Sanction for Violating a Rule 16(f) Order*

Fed. R. Civ. P. 16(f)(1) authorizes a district judge to issue "any just orders," including those authorized by Rule 37(b)(2)(A)(ii)-(vii), if a party or its attorney: "(A) fails to appear at a scheduling or other pretrial conference; (B) is substantially unprepared to participate—or does not participate in good faith—in the conference; or (C) fails to obey a scheduling or other pretrial order."

Rule 37(b)(2)(A) provides in full:

*If a party or a party's officer, director, or managing agent—or a witness designated under Rule 30(b)(6) or 31(a)(4)—fails to obey an order to provide or permit discovery, including an order under Rule 26(f), 35, or 37(a), the court where the action is pending may issue further just orders. They may include the following: (i) directing that the matters embraced in the*

*order or other designated facts be taken as established for purposes of the action, as the prevailing party claims; (ii) prohibiting the disobedient party from supporting or opposing designated claims or defenses, or from introducing designated matters in evidence; (iii) striking pleadings in whole or in part; (iv) staying further proceedings until the order is obeyed; (v) dismissing the action or proceeding in whole or in part; (vi) rendering a default judgment against the disobedient party; or (vii) treating as contempt of court the failure to obey any order except an order to submit to a physical or mental examination.”*

Rule 37(b)(2)(C) adds:

*Instead of or in addition to the orders above, the court must order the disobedient party, the attorney advising that party, or both to pay the reasonable expenses, including attorney’s fees, caused by the failure, unless the failure was substantially justified or other circumstances make an award of expenses unjust.*

There is a mandatory component to Rule 16(f) regarding attorneys’ fees. Rule 16(f)(2) provides:

*Instead of or in addition to any other sanction, the court must order the party, its attorney, or both to pay the reasonable expenses — including attorney’s fees — incurred because of any noncompliance with this rule, unless the noncompliance was substantially justified or other circumstances make an award of expenses unjust.*

*Tracinda Corp. v. DaimlerChrysler AG et al.*, 502 F.3d 212 (3rd Cir. 2007) was not an e-discovery case but held that the late production of a witness’s notes (they were discovered and produced during trial) was sanctionable (\$556,061 was awarded for the costs associated with adjournment of the trial and investigating why the notes were not earlier produced) under Rule 16(f). The Third Circuit explained, “if a litigant knows that even inadvertent failure to produce relevant documents may result in a sanction when the existence of the documents is discovered during trial, the litigant may exercise more care in ensuring that all relevant documents are produced.” *Id.* at 243.

Inadvertence does occur and at times responsive documents are found late and then produced. That result may not be a product of the quality of the search software or the search. Nonetheless, lawyers generally, but particularly lawyers advancing predictive coding as their client’s search tool, will want to be careful that they account for a case like this in the drafting of the Rule 16 Order.

## **CONCLUSION**

In the paper world, many lawyers kept a “hot documents” file. These documents might represent the 10-20 key documents in the case. They were identified through a manual review and were recognized by trial counsel as significant either to the prosecution or defense of a claim.

When information is stored electronically, it might be harder to find those hot documents. But lawyers who will be trying the case will develop their themes and they will be able to identify the type of document that would constitute a hot document if it could be located. Using conventional keyword or computer-assisted search approaches, where the goal is a search for the truth, one will want to find the electronically stored information most relevant to the prosecution or defense of a claim, leaving to the trier of fact the task of

determining whether, in fact, the handful of documents that really matter in a trial will be outcome determinative for one side or the other.

Data-rich parties are desperate for relief from what can be extraordinary costs of production of electronically stored information. While rulemaking may provide relief without compromising the effective administration of justice, search technology, which moves faster than rulemaking, has responded to this need. Predictive coding and other forms of computer-assisted review are here to stay and, in time, may become sufficiently commoditized that lawyers will be able to implement them without the need for consultants, or for judges even to have to consider approving their use.

The Model Rules are going to be modified. Good lawyers will learn how to make new technology work for their clients in competent, cost-effective ways. They will also have to learn when cooperation begins to intrude on client confidentiality and work product and react appropriately.

The Federal Rules of Civil Procedure may in an even more formal way codify the importance of cooperation among counsel to make e-discovery both faster and less expensive without compromising justice or a lawyer's duty of diligence to a client.

As with all matters involving ethics and professional responsibility, lawyers are well advised to remember these rules when dealing with issues of search technology:

- Stay abreast of evolving search technology. Or make sure you have someone helping you who is doing so.
- Know the law and your obligations under the Federal Rules and act appropriately.
- Know what rules of professional conduct apply.
- Think before you act.
- Consult before you commit.
- Get good advice when appropriate to do so.
- Speak up at the right time when appropriate to do so.

Most of all, be reasonable. In the end, it is reasonableness by courts, clients, and counsel that will put parties on the best path to proportional search technology solutions—solutions that are just, speedy and, at least, relative to the matter in controversy, inexpensive.

/jmb

## APPENDIX I: REPORTED CASES WHERE JUDGES HAVE ENDORSED THE COOPERATION PROCLAMATION

- *Da Silva Moore v. Publicis Groupe & MSL Group*, 2012 U.S. Dist. LEXIS 23350, \*8 (S.D.N.Y. Feb. 24, 2012) (In explaining how a judge might evaluate predictive coding as a search tool, Judge Peck quoted from an article he had written: “Of course, the best approach to the use of computer-assisted coding is to follow the Sedona Cooperation Proclamation model. Advise opposing counsel that you plan to use computer-assisted coding and seek agreement; if you cannot, consider whether to abandon predictive coding for that case or go to the court for advance approval.”)
- *In re Facebook PPC Advertising Litigation*, 2011 WL 1324516, \*1-2 (N.D. Cal. Apr. 6, 2011) (Emphasizing that communication among counsel is crucial to the success of electronic discovery and citing the Sedona Conference on the importance of cooperation, the magistrate judge ordered the parties to agree on an ESI protocol that addresses formats of production, and to meet and confer to discuss search terms that had been and would be used by Facebook.)
- *American Federation of State County & Municipal Employees v. Ortho-McNeil-Janssen Pharmaceuticals, Inc.*, 2010 U.S. Dist. LEXIS 135371, \*15 (E.D. Pa. Dec. 21, 2010) (Citing the Proclamation, the district court told the parties “to immediately meet and confer in good faith to cooperatively and independently resolve these disputes to the extent possible,” adding it would give parties “ample time” to do so. “If the Parties are unable to resolve these matters within the time allotted, the Court may require them to submit their discovery processes to a special master, with costs to be borne by the parties.”)
- *DeGeer v. Gillis*, 755 F. Supp. 2d 909, 918, 930-31 (N.D. Ill. 2010) (After endorsing the Proclamation, Magistrate Judge Nolan lamented the consequence of the absence of prior cooperation: “This case demonstrates the importance of candid, meaningful discussion of ESI at the outset of the case, including discovery of ESI from non-parties. Had that been done, the parties should have been able to avoid the issuance of multiple subpoenas to Huron. After service of Defendants’ subpoena, Huron and Defendants should have collaborated on the use of particular search terms and the data custodians to be searched in advance of Huron’s searches. Counsel are ordered to confer in person (not via email, letters, or phone) to establish reasonable limits on the scope of Huron’s future ESI production, including restricting the searches to certain key data custodians and agreeing on a narrow list of search terms and date ranges. Counsel are on notice that going forward the Court expects them to genuinely confer in good faith and make reasonable efforts to work together and compromise on discovery issues whenever possible.”)<sup>89</sup>
- *Home Design Services, Inc. v. Trumble*, 2010 U.S. Dist. LEXIS 46723 (D. Colo. Apr. 9, 2010) (Citing the Cooperation Proclamation, the magistrate judge explained: “Counsel’s case management responsibilities should not be seen as antithetical to their role as advocate.”)

---

<sup>89</sup> A month before her decision in *DeGeer*, Magistrate Judge Nolan admonished counsel in *Tamburo v. Sworkin*, 2010 U.S. Dist. LEXIS 121510, \*10-11 (N.D. Ill. Nov. 17, 2010 (footnote and citation omitted)): “The parties are expected to be familiar with the Case Management Procedures regarding discovery on the Court’s website, the Seventh Circuit’s Electronic Discovery Pilot Program’s Principles Relating to the Discovery of Electronically Stored Information, and the Sedona Conference Cooperation Proclamation, available at [www.thesedonaconference.org](http://www.thesedonaconference.org). The parties are ordered to actively engage in cooperative discussions to facilitate a logical discovery flow. For example, to the extent that the parties have not completed their initial disclosures pursuant to Rule 26(a), or if their initial disclosures require updating, the parties should focus their efforts on completing their Rule 26(a) requirement before proceeding to other discovery requests.”

- *Cartel Asset Management v. Ocwen Financial Corp.*, 2010 U.S. Dist. LEXIS 17857, \*40 (D. Colo. Feb. 8, 2010) (The magistrate judge, citing the Cooperation Proclamation, lamented the conduct of counsel: “The court is left with the impression that counsel are searching for discovery disputes, rather than working cooperatively to avoid or defuse those disagreements. Given the talented attorneys involved in this case, that development is regrettable.”)<sup>90</sup>
- *Building Erection Services Co., L.C. v. American Buildings Co.*, 2010 U.S. Dist. LEXIS 2625, \*5-6 (D. Kan. Jan. 13, 2010) (In a dispute over setting deposition dates, the magistrate judge told counsel: “To help counsel understand their obligations, counsel are directed to read the Sedona Conference Cooperation Proclamation, which this Court has previously endorsed.”)
- *Securities and Exchange Commission v. Collins & Aikman*, 256 F.R.D. 403, 415 (S.D.N.Y. 2009) (The SEC’s lack of cooperation prompted the district court to draw the parties’ attention to the Cooperation Proclamation, which “urges parties to work in a cooperative rather than an adversarial manner to resolve discovery issues in order to stem the ‘rising monetary costs’ of discovery disputes.”)
- *Dunkin’ Donuts Franchised Restaurants LLC v. Grand Central Donuts, Inc.*, 2009 U.S. Dist. LEXIS 52261, \*14-15 (E.D.N.Y. June 19, 2009) (Citing the meet-and-confer provisions of Rule 26(f) and the Sedona Cooperation Proclamation in connection with the need to reduce the “rising costs” of discovery disputes, the magistrate judge ordered defendants to work out their differences.)<sup>91</sup>
- *Covad Communications Co. v. Revonet, Inc.*, 258 F.R.D. 5, 14 (D.D.C. 2009) (Referencing the Cooperation Proclamation, the magistrate judge said that a failure of a party to respond to an invitation to propose search terms is not the kind of “collaboration and cooperation that underlies the hope that the courts can, with the sincere assistance of the parties, manage e-discovery efficiently and with the least expense possible.”)
- *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, 2009 U.S. Dist. LEXIS 22903, \*136 (S.D.N.Y. Mar. 19, 2009) (The magistrate judge admonished the Bar: “[T]he best solution in the entire area of electronic discovery is cooperation among counsel. This Court strongly endorses The Sedona Conference Cooperation Proclamation.”)
- *In re DirecTech Southwest, Inc., Fair Labor Standards Act Litigation*, 2009 U.S. Dist. LEXIS 69142 (E.D. La. Aug. 7, 2009) (The magistrate judge echoes the statements made in *William A. Gross*.)
- *Ford Motor Co. v. Edgewood Props., Inc.*, 257 F.R.D. 418, 426 (D.N.J. 2009) (In denying a motion to compel reproduction of electronically stored information, the magistrate judge, citing *William A. Gross*, admonished counsel: “It is beyond cavil that this entire problem could have been avoided had there been an *explicit* agreement between the parties as to production, but as that ship has sailed, it is without question unduly burdensome to a party months after production to require that party to reconstitute their *entire* production to appease a late objection.” (Emphasis in original.)

---

<sup>90</sup> The judge quoted from the Cooperation Proclamation: “[T]he *Cooperation Proclamation* correctly recognizes that while counsel are ‘retained to be zealous advocates for their clients, they bear a professional obligation to conduct discovery in a diligent and candid manner. . . . Cooperation does not conflict with the advancement of their clients’ interests—it enhances them. Only when lawyers confuse *advocacy* with *adversarial conduct* are these twin duties in conflict.” 2010 U.S. Dist. LEXIS 17857, at \*40-41 (emphasis in original).

<sup>91</sup> This is another case involving a dispute over search terms: “Defendants’ proposed search can be narrowed temporally and the scope of the search terms sought tailored to each employee, since some employees may have knowledge of only issues relevant to one set of counterclaims but not the other. The defendants must provide Dunkin with a list of the employees or former employees whose emails they want searched and the specific search terms to be used for each individual depending on whether they were likely to be involved with issues relating to the termination of the franchise agreement or the performance of the store development agreement.” 2009 U.S. Dist. LEXIS 52261 at \*16.

- *Capitol Records, Inc. v. MP3tunes, LLC*, 2009 WL 2568431 (S.D.N.Y. Aug. 13, 2009) (also citing *William A. Gross*, the magistrate judge was critical of counsel’s unilateral decision on search terms for electronically stored information: “MP3tunes’ unilateral decision regarding its search reflected a failure to heed Magistrate Judge Andrew Peck’s recent ‘wake-up call’ regarding the need for cooperation concerning e-discovery.”)
- *Gipson v. Southwestern Bell Tel. Co.*, 2008 U.S. Dist. LEXIS 103822, \*5-6 (D. Kan. Dec. 23, 2008) (Where 115 motions had been filed in less than 12 months after filing of the complaint, the same magistrate judge ordered counsel to read the Cooperation Proclamation and then to attempt to agree on discovery issues or a Special Master would be appointed.)<sup>92</sup>
- *Mancia v. Mayflower Textile Servs Co.*, 253 F.R.D. 354, 357-60 (D. Md. 2008) (Endorsing the Cooperation Proclamation, Judge Grimm wrote: “It cannot seriously be disputed that compliance with the ‘spirit and purposes’ of these discovery rules requires cooperation by counsel to identify and fulfill legitimate discovery needs, yet avoid seeking discovery the cost and burden of which is disproportionately large to what is at stake in the litigation. Counsel cannot ‘behave responsively’ during discovery unless they do both, which requires cooperation rather than contrariety, communication rather than confrontation.”)

---

<sup>92</sup> See *Evolution, Inc. v. The Suntrust Bank*, 2004 U.S. Dist. LEXIS 20490, \*18 (D. Kan. Sept. 29, 2004) (where the same magistrate judge had to appoint a Special Master to resolve e-discovery issues and ultimately required the defendant to pay 70% of the Special Master’s \$52,000 fee because of its “representations and lack of cooperation”).

## **APPENDIX II: EXCERPTS FROM *ZUBULAKE V*, 229 F.R.D. 422, 432-434 (S.D.N.Y. 2004) ON OUTSIDE COUNSEL'S DUTY TO MONITOR PRODUCTION**

A party's discovery obligations do not end with the implementation of a "litigation hold"—to the contrary, that's only the beginning. Counsel must oversee compliance with the litigation hold, monitoring the party's efforts to retain and produce the relevant documents. Proper communication between a party and her lawyer will ensure (1) that all relevant information (or at least all sources of relevant information) is discovered, (2) that relevant information is retained on a continuing basis; and (3) that relevant nonprivileged material is produced to the opposing party.

\*\*\*

In short, it is *not* sufficient to notify all employees of a litigation hold and expect that the party will then retain and produce all relevant information. Counsel must take affirmative steps to monitor compliance so that all sources of discoverable information are identified and searched. This is not to say that counsel will necessarily succeed in locating all such sources, or that the later discovery of new sources is evidence of a lack of effort. But counsel and client must take *some reasonable steps* to see that sources of relevant information are located.

The tricky question is what that continuing duty entails. What must a lawyer do to make certain that relevant information—especially electronic information—is being retained? Is it sufficient if she periodically resends her initial "litigation hold" instructions? What if she communicates with the party's information technology personnel? Must she make occasional on-site inspections?

Above all, the requirement must be reasonable. A lawyer cannot be obliged to monitor her client like a parent watching a child. At some point, the client must bear responsibility for a failure to preserve. At the same time, counsel is more conscious of the contours of the preservation obligation; a party cannot reasonably be trusted to receive the "litigation hold" instruction once and to fully comply with it without the active supervision of counsel.

*First*, counsel must issue a "litigation hold" at the outset of litigation or whenever litigation is reasonably anticipated. The litigation hold should be periodically reissued so that new employees are aware of it, and so that it is fresh in the minds of all employees.

*Second*, counsel should communicate directly with the "key players" in the litigation, *i.e.*, the people identified in a party's initial disclosure and any subsequent supplementation thereto. Because these "key players" are the "employees likely to have relevant information," it is particularly important that the preservation duty be communicated clearly to them. As with the litigation hold, the key players should be periodically reminded that the preservation duty is still in place.

*Finally*, counsel should instruct all employees to produce electronic copies of their relevant active files. Counsel must also make sure that all backup media which the party is required to retain is identified and stored in a safe place. In cases involving a small number of relevant backup tapes, counsel might be advised to take physical possession of backup tapes. In other cases, it might make sense for relevant backup tapes to be segregated and placed in storage. Regardless of what particular arrangement counsel chooses to employ, the point is to separate relevant backup tapes from others. One of the primary reasons that electronic data is lost is ineffective communication with information technology personnel. By taking possession of, or otherwise

safeguarding, all potentially relevant backup tapes, counsel eliminates the possibility that such tapes will be inadvertently recycled.

## ABOUT THE AUTHOR

### John M. Barkett

Mr. Barkett is a partner at the law firm of Shook, Hardy & Bacon L.L.P. in its Miami office. He is a graduate of the University of Notre Dame (B.A. Government, 1972, *summa cum laude*) and the Yale Law School (J.D. 1975) and served as a law clerk to the Honorable David W. Dyer on the old Fifth Circuit Court of Appeals. Mr. Barkett is an adjunct professor of law at the University of Miami School of Law. He is also the recipient of one of the 2011 Burton Awards for Legal Achievement which honors lawyers for distinguished legal writing. In March 2012, the Chief Justice appointed Mr. Barkett to serve on the Advisory Committee for Civil Rules of the Federal Judicial Conference. Mr. Barkett is also a former member of the Council of the ABA Section of Litigation.

Mr. Barkett has, over the years, been a commercial litigator (contract and corporate disputes, employment, trademark, and antitrust), environmental litigator (CERCLA, RCRA, and toxic tort), and, for the past several years, a peacemaker and problem solver, serving as an arbitrator, mediator, facilitator, or allocator in a variety of environmental, commercial, or reinsurance contexts. He is a certified mediator under the rules of the Supreme Court of Florida and the Southern and Middle Districts of Florida and a member of the London Court of International Arbitration and the International Council for Commercial Arbitration, and serves on the AAA and ICDR roster of neutrals, the CPR Institute for Dispute Resolution's "Panel of Distinguished Neutrals," and the National Roster of Environmental Dispute Resolution and Consensus Building Professionals maintained by the U.S. Institute for Environmental Conflict Resolution. He has served or is serving as a neutral in scores of matters involving in the aggregate more than \$4 billion. He has conducted or is conducting commercial domestic and international arbitrations under AAA, LCIA, ICDR, UNCITRAL, and CPR rules and has conducted *ad hoc* arbitrations. In November 2003, he was appointed by the presiding judge to serve as the Special Master to oversee the implementation and enforcement of the 1992 Consent Decree between the United States and the State of Florida relating to the restoration of the Florida Everglades. He also consults with major corporations on the evaluation of legal strategy and risk and conducts independent investigations where such services are needed.

Mr. Barkett has published two books, *E-Discovery: Twenty Questions and Answers*, (Chicago: First Chair Press, 2008) and *The Ethics of E-Discovery* (Chicago: First Chair Press, 2009). Mr. Barkett has also prepared analyses of the Roberts Court the past six years, in addition to a number of other articles on a variety of topics:

- *The Roberts Court 2011-12: The Affordable Care Act and More* (ABA Annual Meeting, Chicago, August 3, 2012)
- *Ethical Challenges on the Horizon: Confidentiality, Competence and Cloud Computing* (ABA-CLE, July 24, 2012)
- *Un-taxing E-Discovery Costs: Section 1920(4) After Race Tire Amer. Inc. and Taniguchi* (June 29, 2012) (<http://www.shb.com/attorneys/BarkettJohn/UntaxingEdiscoveryCosts.pdf>)
- Barkett, *ABA to Tackle Technology Issues in Model Rules at August Meeting*, (<http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202560335059&thepage=3&slreturn=1>). Law Technology News (June 25, 2012)

- *E-Communications: Problems Posed by Privilege, Privacy, and Production* (ABA National Institute on E-Discovery, New York, NY, May 18, 2012)
- *The 7<sup>th</sup> Circuit Pilot Project: What We Might Learn And Why It Matters to Every Litigant in America* (ABA Section of Litigation News Online, December 11, 2011) [http://apps.americanbar.org/litigation/litigationnews/civil\\_procedure/docs/barkett.december11.pdf](http://apps.americanbar.org/litigation/litigationnews/civil_procedure/docs/barkett.december11.pdf)
- *Skinner, Matrixx, Souter, and Posner: Iqbal and Twombly Revisited*, 12 *The Sedona Conference Journal* 69 (2011) (Mr. Barkett received the Burton Award for Legal Achievement for this paper).
- *The Challenge of Electronic Communication, Privilege, Privacy, and Other Myths*, 38 *Litigation Journal* 17 (ABA Section of Litigation, Fall 2011)
- *Avoiding the Cost of International Commercial Arbitration: Is Mediation the Solution?* in Contemporary Issues in International Arbitration and Mediation – The Fordham Papers (Martinus Nijhoff, New York, 2011)
- *The Roberts Court 2010-11: Three Women Justices!* (ABA Annual Meeting, Toronto, August 2011)
- *The Ethics of Web 2.0*, (ACEDS Conference, Hollywood, FL March 2011)
- *The Roberts Court: Year Four, Welcome Justice Sotomayor* (ABA Annual Meeting, San Francisco, August 2010)
- *The Myth of Culture Clash in International Commercial Arbitration* (co-authored with Jan Paulsson), 5 *Florida International University Law Review* 1 (June 2010)
- *Walking the Plank, Looking Over Your Shoulder, Fearing Sharks Are in the Water: E-Discovery in Federal Litigation?* (Duke 2010 Conference, Civil Rules Advisory Committee, May 11, 2010) (<http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Duke%20Materials/Library/John%20Barkett,%20Walking%20the%20Plank.pdf>)
- *Zubulake Revisited: Pension Committee and the Duty to Preserve*, (Feb. 26, 2010) ([http://www.abanet.org/litigation/litigationnews/trial\\_skills/pension-committee-zubulake-ediscovery.html](http://www.abanet.org/litigation/litigationnews/trial_skills/pension-committee-zubulake-ediscovery.html))
- *Draft Reports and Attorney-Expert Communications*, 24 *N.R.E.* (Winter 2010)
- *From Canons to Cannon in A Century of Legal Ethics: Trial Lawyers and the ABA Canons of Professional Ethics* (American Bar Association, Chicago, 2009)
- *The Robert's Court: Three's a Charm* (ABA Annual Meeting, Chicago, August 2009)
- *Cheap Talk? Witness Payments and Conferring with Testify Witnesses*, (ABA Annual Meeting, Chicago, 2009)
- *Burlington Northern: The Super Quake and Its Aftershocks*, 58 *Chemical Waste Lit. Rprt.* 5 (June 2009)
- *Fool's Gold: The Mining of Metadata* (ABA's Third Annual National Institute on E-Discovery, Chicago, May 22, 2009)
- *More on the Ethics of E-Discovery* (ABA's Third Annual National Institute on E-Discovery, Chicago, May 22, 2009)
- *Production of Electronically Stored Information in Arbitration: Sufficiency of the IBA Rules in Electronic Disclosure in International Arbitration* (JurisNet LLC, New York, September 2008)
- *The Robert's Court: The Terrible Two's or Childhood Bliss?* (ABA Annual Meeting, New York, August 2008)
- *Orphan Shares*, 23 *NRE* 46 (Summer 2008)
- *Tipping The Scales of Justice: The Rise of ADR*, 22 *NRE* 40 (Spring 2008)
- *Tattletales or Crimestoppers: Disclosure Ethics Under Model Rules 1.6 and 1.13*, (ABA Annual Meeting, Atlanta, August 7, 2004 and, in an updated version, ABA Tort and Insurance Practice Section Spring CLE Meeting, Phoenix, April 11, 2008)

- *E-Discovery For Arbitrators*, 1 Dispute Resolution International Journal 129, International Bar Association (Dec. 2007)
- *The Roberts Court: Where It's Been and Where It's Going* (ABA Annual Meeting, San Francisco, August, 2007)
- *Help Has Arrived...Sort Of: The New E-Discovery Rules*, ABA Section of Litigation Annual Meeting, San Antonio (2007)
- *Refresher Ethics: Conflicts of Interest*, (January 2007 ABA Section of Litigation Joint Environmental, Products Liability, and Mass Torts CLE program)
- *Help Is On The Way...Sort Of: How the Civil Rules Advisory Committee Hopes to Fill the E-Discovery Void*, ABA Section of Litigation Annual Meeting, Los Angeles (2006)
- *The Battle For Bytes: New Rule 26, e-Discovery*, Section of Litigation (February 2006)
- *Forward to the Past: The Aftermath of Aviall*, 20 N.R.E. 27 (Winter 2006)
- *The Prelitigation Duty to Preserve: Lookout!* ABA Annual Meeting, Chicago, (2005)
- *The MJP Maze: Avoiding the Unauthorized Practice of Law* (2005 ABA Section of Litigation Annual Conference)
- *Bytes, Bits and Bucks: Cost-Shifting and Sanctions in E-Discovery*, ABA Section of Litigation Annual Meeting (2004) and 71 Def. Couns. J. 334 (2004)
- *The CERCLA Limitations Puzzle*, 19 N.R.E. 70 (Fall, 2004)
- *If Terror Reigns, Will Torts Follow?* 9 Widener Law Symposium 485 (2003)

Mr. Barkett is also the author of Ethical Issues in Environmental Dispute Resolution, a chapter in the ABA publication, *Environmental Dispute Resolution, An Anthology of Practical Experience* (July 2002) and the editor and one of the authors of the ABA Section of Litigation's Monograph, *Ex Parte Contacts with Former Employees* (Environmental Litigation Committee, October 2002).

Mr. Barkett is a Fellow of the American College of Environmental Lawyers. At the University of Miami Law School, Mr. Barkett teaches two courses, "Environmental Litigation" and "E-Discovery."

Mr. Barkett has been recognized in the areas of alternative dispute resolution or environmental law in a number of lawyer-recognition publications, including Who's Who Legal (International Bar Association) (since 2005); Best Lawyers in America (National Law Journal) (since 2005); Legal Elite (since 2004), (Florida Trend), Florida Super Lawyers (since 2008), and Chambers USA America's Leading Lawyers (since 2004). Mr. Barkett can be reached at [jbarkett@shb.com](mailto:jbarkett@shb.com).