

# Richmond Journal of Law and Technology

---

Volume 18 | Issue 3

Article 4

---

2012

## Technology-Assisted Document Review: Is it Defensible?

William W. Belt

Dennis R. Kiker

Daryl E. Shetterly

Follow this and additional works at: <http://scholarship.richmond.edu/jolt>

 Part of the [Computer Law Commons](#)

---

### Recommended Citation

William W. Belt, Dennis R. Kiker & Daryl E. Shetterly, *Technology-Assisted Document Review: Is it Defensible?*, 18 Rich. J.L. & Tech 10 (2012).

Available at: <http://scholarship.richmond.edu/jolt/vol18/iss3/4>

This Article is brought to you for free and open access by UR Scholarship Repository. It has been accepted for inclusion in Richmond Journal of Law and Technology by an authorized administrator of UR Scholarship Repository. For more information, please contact [scholarshiprepository@richmond.edu](mailto:scholarshiprepository@richmond.edu).

## **TECHNOLOGY-ASSISTED DOCUMENT REVIEW: IS IT DEFENSIBLE?**

By William W. Belt, Dennis R. Kiker and Daryl E. Shetterly\*

Cite as: William W. Belt, Dennis R. Kiker & Daryl E. Shetterly, *Technology-Assisted Document Review: Is It Defensible?*, XVIII RICH. J. L. & TECH 10 (2012), <http://jolt.richmond.edu/v18i3/article10.pdf>.

### **I. INTRODUCTION**

[1] Technology has changed the way we communicate and, in so doing, has changed the discovery phase of litigation. Parties must sift through ever-growing data volumes to find relevant material, significantly increasing time and cost requirements. Technology has also changed the way attorneys meet discovery demands. New technologies like “machine learning” and “predictive coding”<sup>1</sup> give lawyers important new tools to manage the growing volume of electronically stored information (“ESI”).<sup>2</sup>

---

\* William W. Belt is the leader of LeClairRyan’s electronic discovery practice group. Dennis R. Kiker and Daryl E. Shetterly are partners in the electronic discovery practice group.

<sup>1</sup> See Jason. R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on ‘Information Inflation’ and Current Issues in E-Discovery Search*, 17 RICH. J. L. & TECH. 9, at 25-26 (2011), <http://jolt.richmond.edu/v17i3/article9.pdf> (explaining terms like “machine learning” and “predictive coding” are just two of many terms used to refer to technology-assisted review). The technology and applications that we call “technology-assisted review,” “predictive coding” and “machine learning” continue to change in important ways, and are therefore difficult to accurately define. In this article, “technology-assisted review” is a family of technologies and applications that receive input such as coding decisions from humans for a subset of documents, and use that input to help categorize, “predictively” code, or rank the remaining documents in the set. *Id.*

<sup>2</sup> See *id.*; see also FED. R. CIV. P. 26(b)(2)(B) (providing special rules for ESI).

At the same time, court decisions have sent “wake-up call[s]”<sup>3</sup> warning attorneys that deploying technology without appropriate safeguards may be foolishly rushing in “where angels fear to tread.”<sup>4</sup>

[2] There was a time when clients sent their lawyers a file folder or box of paper containing the documents relevant to litigation. Thanks to the proliferation of email and other ESI,<sup>5</sup> documents now more commonly arrive on a hard drive, and that hard drive likely contains gigabytes or terabytes of data which, if printed, would fill the law firm’s halls with boxes of paper.<sup>6</sup> At first, the shift from reviewing and analyzing data in paper format to electronic format did little to change the document review process.<sup>7</sup> Attorneys sat in front of computer screens and looked through email inboxes chronologically, similar to the way they previously would

---

<sup>3</sup> William A. Gross Constr. Assocs. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 134 (S.D.N.Y. 2009).

<sup>4</sup> United States v. O’Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008).

<sup>5</sup> John Gantz & David Reinsel, *Extracting Value from Chaos*, EMC CORP. (June 2011), <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf> (pointing out that the world’s information is “more than doubling every two years.”) In 2011 the world will create a staggering 1.8 zettabytes of information. *Id.*

<sup>6</sup> The Sedona Conference, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189, 192 n.2 (2007) [hereinafter *Sedona Search Commentary*] (noting that “[o]ne gigabyte of electronic information can generate approximately 70,000-80,000 of text pages, or 35 to 40 banker’s boxes of documents (at 2,000 pages per box). Thus, a 100-gigabyte storage device (e.g., a personal computer hard drive), theoretically, could hold as much as the equivalent of 3,500 to 4,000 banker’s boxes of documents. By contrast, in 1990, a typical personal computer held just 200 megabytes of data - 1/500 the capacity of a typical hard drive today. Even if only 10% of a computer’s available capacity today contains useful or “useable” information (as distinguished from application programs, operating systems, utilities, etc.), attorneys still would need to consider and potentially review 700,000 to 800,000 pages per each device.”).

<sup>7</sup> *See id.* at 193.

have read through a box of paper.<sup>8</sup> e-Discovery technology has continually evolved to offer new tools and solutions. Now counsel has a myriad of tools available to assist in locating and reviewing relevant documents. With these technological advancements, the need has grown for technological expertise. Attorneys must understand the tools they deploy and how they fit in the discovery process. For most trial lawyers, the need to understand new technologies – both the technologies clients use to communicate and the technologies attorneys may use in discovery – can create daunting challenges.

[3] In response to evolving technology, the people and processes used to solve electronic discovery problems have continually changed since the earliest days of electronic discovery. In the few short years since electronic discovery emerged as an industry, litigants and attorneys have felt the “future shock” of accelerating change.<sup>9</sup> Technology-assisted review is yet another jolt to attorneys—a technology with the potential to change the methods we use to comply with our electronic discovery obligations.

[4] Attorneys have been hesitant to adopt each succeeding generation of document review technology, including technology-assisted review.<sup>10</sup> There are likely several reasons for this hesitancy.<sup>11</sup> One reason is the

---

<sup>8</sup> *See id.*

<sup>9</sup> *See* ALVIN TOFFLER, *FUTURE SHOCK* 4 (1970) (explaining that Toffler “coined the term ‘future shock’ to describe the shattering stress and disorientation that we induce in individuals by subjecting them to too much change in too short a time.”).

<sup>10</sup> *See* Rachel K. Alexander, *E-Discovery Practice, Theory, and Precedent: Finding the Right Pond, Lure, and Lines Without Going on a Fishing Expedition*, 56 S.D. L. REV. 25, 27 (2011).

<sup>11</sup> We do not argue that technology-assisted review is the right tool for every case. Examples of document populations that may not be good candidates for technology-assisted review include; small document sets, document sets containing non-standard document types and document sets with a high percentage of paper documents or image files with text generated by optical character recognition software.

cautionary messages sent by court rulings like *O'Keefe* and *William Gross*.<sup>12</sup> Those two decisions relate to keyword searching, which has for some time been considered safe territory.<sup>13</sup> Moreover, technology-assisted review requires legal, technological and business process sophistication to effectively incorporate the technology into a large-scale discovery project.<sup>14</sup> In other words, attorneys must understand how to integrate technology-assisted review with the human component of document review.<sup>15</sup> Though the same is true for other methods for

---

<sup>12</sup> See generally *United States v. O'Keefe*, 537 F. Supp. 2d 14, 24 (D.D.C. 2008); *William A. Gross Constr. Assocs. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134, 134 (S.D.N.Y. 2009).

<sup>13</sup> See generally *id.*

<sup>14</sup> See Patrick Oot, Anne Kershaw, & Herbert L. Roitblat, *Mandating Reasonableness in a Reasonable Inquiry*, 87 DENV. U. L. REV. 533, 534-35 (2010).

<sup>15</sup> See, e.g., Ralph Losey, *Bottom Line Driven Proportional Review*, E-DISCOVERY TEAM (Jan. 15, 2012), <http://e-discoveryteam.com/2012/01/15/bottom-line-driven-proportional-review/> (“[Y]ou cannot just dispense with final manual review [...] we are not going to turn that over to the Borg anytime soon. I’ve asked around and no law firms do that now. No experts advocate that approach either, even the most extreme advocates for automation (of which I’m one) [...] You use predictive coding to speed up the final manual review to be sure, but only a fool (or con artist trying to get at a producing parties [sic] secrets) trusts coding software today without human verification.”); see also MAURA R. GROSSMAN, CONOR R. CROWLEY & JOE LOOBY, TREC, REFLECTIONS OF THE TOPIC AUTHORITIES (2008), available at <http://trec-legal.umiacs.umd.edu/other/TArefections2008.doc> (explaining “how ‘responsiveness’ is defined is often dependent on numerous subjective determinations involving, among other things, the nature of the risk posed by production, the party requesting the information, the willingness of the producing party to face a challenge for underproduction, and the level of knowledge that the producing party has about the matter at a particular point in time. Lawyers can and do draw these lines differently for different types of opponents, on different matters, and at different times on the same matter. This makes it exceedingly difficult to establish a ‘gold standard’ against which to measure relevance/responsiveness and explains why document review cannot be completely automated.”); Herbert L. Roitblat, Anne Kershaw & Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. AM. SOC’Y INFO. SCI. 1, 8 (2009), available at <http://www.clearwellsystems.com/e-discovery-blog/wp-content/uploads/2010/12/man-v-comp-doc-review.pdf> (“Discovery cannot be wholly automated, not for the reason that it

facilitating document review, such as keyword searching, the complexity of the technology and importance of the process are new territory for most lawyers.

[5] Most importantly, uncertainty remains as to whether the use of technology-assisted review tools is legally defensible.<sup>16</sup> Though intellectual debate challenges the efficacy of keyword searching, it is generally-accepted and widely used. Technology-assisted review is not as of yet. Judge Peck argues that counsel may be waiting for an opinion stating that technology-assisted review is, or is not, a reasonable means of identifying relevant information.<sup>17</sup> Anticipating that day, and in the interest of furthering the academic discussion around technology-assisted review, included herein is a legal brief that supports the use of technology-assisted review in a hypothetical case.

[6] In this hypothetical, the producing defendant faces a motion to compel after using technology-assisted review to exclude from review a subset of documents that technology has “predictively coded” as not likely to contain relevant information. During the meet and confer process, plaintiff objected to using the technology and insisted that the producing

---

involves so-called subjective judgment, but because ultimately attorneys and parties in the case have to know what the data are about. They have to formulate and respond to arguments and develop a strategy for winning the case. They have to understand the evidence that they have available and be able to refute contrary evidence. All of this takes knowledge of the case, the law, and much more.”).

<sup>16</sup> See, e.g., Andrew Peck, *Search, Forward*, LAW TECH. NEWS (Oct. 1, 2011), <http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202516530534>. See also, *Da Silva Moore v. Publicis Groupe, et al.*, As of the date of this writing, no order is available, but a transcript of a hearing before Judge Peck addressing the technology is available at <http://www.law.com/jsp/lawtechnologynews/PubArticleFriendlyLTN.jsp?id=1202542221714&slreturn=1>.

<sup>17</sup> *Id.* (“Perhaps they are looking for an opinion concluding that: ‘It is the opinion of this court that the use of predictive coding is a proper and acceptable means of conducting searches under the Federal Rules of Civil Procedure, and furthermore that the software provided for this purpose by [insert name of your favorite vendor] is the software of choice in this court.’ If so, it will be a long wait.”).

party review all documents, including those predictively coded as not likely to be relevant. The defendant used the technology without obtaining plaintiff's consent, and plaintiff later obtained relevant documents from a third party that were excluded from production by the technology. Plaintiff filed a motion to compel defendant to review all of the documents that had been excluded through technology-assisted review, and defendant filed this brief in response. The brief in this hypothetical case is offered to provide a starting point from which to discuss the issues in the context of a court motion. The brief is written from the perspective of the technology proponent; however, in an actual case, corresponding briefs opposing the technology would precede and follow the response brief. There is not space here to include the opponent's arguments. In addition, the provided hypothetical brief does not address in detail the complex safeguards the courts require when counsel deploys technology in the discovery process. The sampling process, for example, may involve a statistical analysis better suited to a separate study. The arguments are based on federal law, though they should prove applicable in many state courts as well.<sup>18</sup>

---

<sup>18</sup> In addition to the 2006 amendments to the Federal Rules of Civil Procedure, many states have added language to their statutes or rules to accommodate electronic discovery. *See Current Listing of States that Have Enacted e-Discovery Rules*, ELECTRONIC DISCOVERY L., <http://www.ediscoverylaw.com/promo/state-district-court-rules/> (last visited Feb. 1, 2012).

**II. HYPOTHETICAL BRIEF**IN THE UNITED STATES DISTRICT COURT  
FOR THE EASTERN DISTRICT OF MOUNTAIN STATE

|  |                                      |                            |
|--|--------------------------------------|----------------------------|
| Plaintiff, Inc.,<br><br>vs.<br><br>Defendant, Inc. | Plaintiff,<br><br><br><br>Defendant. | Civil Action No. 000-00001 |
|--|--------------------------------------|----------------------------|

DEFENDANT’S RESPONSE TO PLAINTIFF’S MOTION TO  
COMPEL REVIEW AND PRODUCTION OF DOCUMENTS  
EXCLUDED FROM REVIEW BY TECHNOLOGY-ASSISTED  
REVIEW METHODOLOGY AND MOTION FOR PROTECTIVE  
ORDER<sup>19</sup>

Defendant, Defendant, Inc. (“Defendant”) submits the following combined response to Plaintiff’s Motion to Compel and Defendant’s Motion for Protective Order. The Court should deny Plaintiff’s motion because the technology-assisted review process used by Defendant in this case was reasonable and satisfies Defendant’s discovery obligations under

---

<sup>19</sup> As this “brief” is presented solely for academic discussion, the format and style may not be appropriate for a brief filed with a court. *See* discussion *supra* p. 4.



the Federal Rules of Civil Procedure. Defendant combined the human review of 200,000 responsive and privileged documents with the technology-assisted review of 800,000 documents categorized by the technology<sup>20</sup> as “not relevant.” Defendant’s Response to Plaintiff’s Motion to Compel at ¶¶ 8-10. Defendant respectfully submits that the process was reasonable and achieves the underlying goal of the Federal Rules of Civil Procedure to “secure the just, speedy, and inexpensive determination of every action and proceeding,” by ensuring that the actions taken to identify and produce relevant information do not “outweigh[] its likely benefit, considering the needs of the case, the amount in controversy, the parties’ resources, the importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.” Fed. R. Civ. P. 1, 26(b)(2)(C).

## I. STATEMENT OF FACTS

1. Defendant and Plaintiff identified 20 custodians in Defendant’s employ that were most likely to have information and

---

<sup>20</sup> Defendant’s attorneys also reviewed and coded a statistical sample of randomly selected documents to allow the technology to categorize the entire set. *See infra* Part I, at 8.

documents relevant to the underlying matter. The parties also agreed upon a relevant date range for discovery. The details of the parties' agreement is contained in the ESI Protocol, attached as Exhibit A.<sup>21</sup>

2. Defendant collected documents in accordance with the ESI Protocol, which provides that Defendant would collect all e-mail and active files associated with the identified custodians. Defendant then engaged a third party provider, Vendor, Inc., to process and host the documents for review. After processing and de-duplication, 1 million unique documents were loaded into the review application for attorney review.

3. Recognizing the significant cost associated with reviewing each and every one of the million documents loaded into the review

---

<sup>21</sup> Since this is a brief in a hypothetical, there are no exhibits attached. Though the ESI protocol depends on the circumstances of each case, a number of resources may serve as a starting point for an ESI protocol. *See, e.g., Suggested Protocol for Discovery of Electronically Stored Information ('ESI')*, U.S. DISTRICT CT. FOR THE DISTRICT OF MD., available at <http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf> (last visited Feb. 4, 2012); *An E-Discovery Model Order*, U.S. CT. APPEALS FOR THE FED. CIRCUIT, available at [http://www.cafc.uscourts.gov/images/stories/announcements/Ediscovery\\_Model\\_Order.pdf](http://www.cafc.uscourts.gov/images/stories/announcements/Ediscovery_Model_Order.pdf) (last visited Feb. 4, 2012); *Default Standard for the Discovery of Electronically Stored Information ("E-Discovery")*, U.S. DISTRICT CT. FOR THE N. DISTRICT OF OHIO, available at [http://www.ohnd.uscourts.gov/assets/Rules\\_and\\_Orders/Local\\_Civil\\_Rules/CoverSheet.htm](http://www.ohnd.uscourts.gov/assets/Rules_and_Orders/Local_Civil_Rules/CoverSheet.htm) (follow "Appendices" tab; then follow "Appendix K" hyperlink) (last visited Feb. 4, 2012).

application, and the likelihood that many of those documents were not relevant to any matter at issue in this case, Defendant proposed to use a technology-assisted review tool<sup>22</sup> to divide the documents into two categories: (1) documents likely to be relevant; and (2) documents likely to be not relevant.<sup>23</sup> Defendant further proposed limiting human review to: (a) an initial set of randomly selected documents that would be reviewed by attorneys so that their coding decisions could be applied to the rest of the data set; (b) the data set that the tool identified as most likely to be relevant; and (c) a random, statistically significant sample of

---

<sup>22</sup> Technology-assisted review is also referred to as machine learning, predictive coding, software assisted review and suggestive coding. See Jim Eidelman & Ron Tienzo, *Predictive Coding & Non-Linear Review: Best Practices and Comparative Analysis*, CATALYST, available at [http://www.catalystsecure.com/Webinars/pdfs/Partner\\_Predictive\\_Coding\\_and\\_Non-Linear%20Review\\_Webinar\\_Dec\\_15\\_2011.pdf](http://www.catalystsecure.com/Webinars/pdfs/Partner_Predictive_Coding_and_Non-Linear%20Review_Webinar_Dec_15_2011.pdf). The term “technology-assisted review” is used throughout this brief to refer to these and other technologies that receive input, such as coding decisions, from humans for a subset of documents, and apply that input to help categorize, “predictively” code, or rank the remaining documents in the set.

<sup>23</sup> This refers to technology-assisted review in the context of filtering data that has been collected from its original environment and indexed for search. Technology-assisted review tools may also be used to filter data in its native environment, but that application is limited to instances where the technology is deployed behind the firewall. Cf. Gordon V. Cormack & Mona Mojdeh, *Machine Learning for Information Retrieval: TREC 2009 Web, Relevance Feedback and Legal Tracks*, in THE EIGHTEENTH TEXT RETRIEVAL CONFERENCE (TREC 2009) PROCEEDINGS at 3 (2009), available at <http://trec.nist.gov/pubs/trec18/papers/uwaterloo-cormack.WEB.RF.LEGAL.pdf>.

the documents the technology-assisted review technology identified as “not relevant.”

4. Plaintiff refused to agree to Defendant’s proposed technology-assisted review and sampling processes, and proposed instead that Defendant run several hundred search terms across the entire volume of data and review all documents that contained any of those search terms. The list of Plaintiff’s proposed search terms is attached as Exhibit B.

5. Defendant ran the proposed search terms as requested and discovered that 967,453 of the documents (inclusive of family members)<sup>24</sup> contained one or more of the search terms, which included such common terms as “manufacture” and “quality control.”

6. Under the Scheduling Order, Defendant had 30 days to complete the review and production. The deadline meant there was no

---

<sup>24</sup> The phrase “family member” in the e-Discovery context refers to an attachment. *See, e.g.,* Steve Green, *Document Family Circus*, DISCOVERY IN PRAC. (Apr. 11, 2011), <http://hudsonlegalblog.com/e-discovery/the-document-family-circus.html> (explaining an e-mail and its attachment are generally seen as two separate documents but are often considered part of the same “family” of documents for review and production purposes). While the number of documents that actually contain the search terms requested may be lower, it is often necessary to view a document in context with its attachments to determine privilege and responsiveness.

time to effectively negotiate further keyword limitations or to add Boolean connectors and complete the review on time.<sup>25</sup>

7. Because Defendant felt the search term protocol was ineffective in identifying only relevant documents, and in light of the time constraints and Plaintiff's refusal to agree, Defendant elected to use a technology-assisted review tool to identify the documents in the population most likely to be relevant. The technology-assisted review tool and process, described in detail below, utilized a subset of the search terms proposed by Plaintiff, attached as Exhibit C.

8. The technology-assisted review tool identified approximately 200,000 documents likely to have relevant information.

9. Defendant's attorneys reviewed all of these documents for privilege, confidentiality, and trade secrets, and subsequently produced 149,376 relevant, non-privileged documents.

---

<sup>25</sup> The timing and deadlines inherent in litigation are often overlooked in the discovery process. We include a deadline here to underscore the role timing plays in managing discovery projects. While computers can increase speed, data volumes can offset the advantage and make deadlines more difficult to meet. *See* Dean Gonsowski, *A Look into the Crystal Ball: E-Discovery Predictions and Trends*, ALANET.ORG (July/Aug. 2010), <http://www.alanet.org/publications/issue/julaug10/LM-JulAug10-F1-EDiscovery.pdf>.

10. Defendant then reviewed a random sample of the remaining 800,000 documents and, finding no additional relevant documents, determined with a 95% confidence level that less than 3% of the unreviewed documents were relevant.<sup>26</sup>

11. Defendant produced 149,376 documents on time under the Scheduling Order.

12. Following production, Plaintiff identified and supplemented its production with a relevant e-mail, sent by one of the custodians during the relevant date range, that was produced pursuant to a third-party subpoena to Third Party Corp., but not included in Defendant's document production ("supplemented message"). Defendant was able to locate the supplemented message among the 800,000 documents that were excluded from review. The supplemented message was not in the random sample of documents reviewed as part of the quality control process.

---

<sup>26</sup> The exact number of documents that need to be reviewed to determine confidence level and confidence interval varies with the size of the document population. There are several resources available to identify the number of documents that need to be reviewed to determine the confidence level and confidence interval. *See, e.g., Sample Size Calculator*, CREATIVE RES. SYS., <http://www.surveysystem.com/sscalc.htm#one> (last visited Feb. 1, 2012).

13. Following receipt of the supplemented message, Plaintiff renewed its demand that Defendant undertake a manual, linear review of the remaining 800,000 documents, and, when Defendant refused, filed its Motion to Compel.

## II. OVERVIEW OF THE TECHNOLOGY-ASSISTED DOCUMENT REVIEW PROCESS

### A. Why Do We Need Technology-Assisted Review?

In the past decade, there has been an explosion in the volume of electronic information retained by organizations.<sup>27</sup> As a result, litigators must work with their clients to sift through larger and larger data sets to identify the relevant documents they are required to produce to comply with their obligations under the Federal Rules.<sup>28</sup>

There is generally no obligation for lawyers to look at every document within the organization to determine whether it is relevant to the

---

<sup>27</sup> Gantz, *supra* note 5.

<sup>28</sup> The Federal Rules of Civil Procedure require producing non-privileged documents responsive to requests “reasonably calculated to lead to the discovery of admissible evidence.” FED. R. CIV. P. 26(b)(1).

litigation.<sup>29</sup> Lawyers narrow the inquiry by interviewing employees to identify custodians – employees that have the relevant business records – and identifying data environments that contain relevant electronic and paper documents.<sup>30</sup> They work with information technology groups (hereinafter “IT”) to collect those documents in preparation for production to the opposing party.<sup>31</sup>

---

<sup>29</sup> *Id.*; see also The Sedona Conference, BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION 38 (Jonathon M. Redgrave et al. eds., 2d ed., 2007) [hereinafter SEDONA PRINCIPLES], available at [http://www.thesedonaconference.org/dltForm?did=TSC\\_PRINCP\\_2nd\\_ed\\_607.pdf](http://www.thesedonaconference.org/dltForm?did=TSC_PRINCP_2nd_ed_607.pdf) (organizations should define the search for relevant documents by limiting their search to “electronically stored information from repositories used by key individuals rather than generally searching through the entire organization’s electronic information systems.”).

<sup>30</sup> See Bernd Honsel, Gerald G. Paul & Wolfgang A. Dase, *Representing European Companies in U.S. Litigation: Document handling—Document custodians*, in 1 *Successful Partnering Between Inside and Outside Counsel* § 23:19 (“Document custodians, including the company’s information technology personnel, can be of great importance as litigation unfolds. In particular, if they are long-time employees of the company they may have a wealth of knowledge concerning documents under their supervision, including documents that may be vital to the company’s position in the litigation.”); see generally Daryl Shetterly, *Getting the Most from the Custodian Interview*, THE E-DISCOVERY MYTH (Dec. 19, 2011), <http://e-discovery.com/2011/12/19/getting-the-most-from-the-custodian-interview/>.

<sup>31</sup> See, e.g., *Zubulake v. UBS Warburg LLC*, 229 F.R.D. 422, 432 (S.D.N.Y. 2004).



A decade ago, the documents copied in preparation for production were more likely to contain only the relevant, non-privileged documents.<sup>32</sup> Even if the data set contained other documents, a few lawyers or paralegals could sort through and categorize them efficiently with limited cost.<sup>33</sup> As data volumes grew, so too did the volume of irrelevant documents comingled with the documents collected from the client.<sup>34</sup> Given the large data volumes now collected in many cases, the document review phase (separating the relevant documents from the irrelevant documents, and identifying documents to be withheld or redacted and logged as privileged) of an electronic discovery project is often the most expensive part.<sup>35</sup>

Some of the common criteria or tools used to limit the volume of documents that need to be reviewed by humans include limiting document

---

<sup>32</sup> Craig Ball, *The Plaintiff's Practical Guide to E-Discovery, Part I*, CRAIG D. BALL P.C., 2 (2005), <http://www.craigball.com/EDD-The%20Practical%20Plaintiffs%20Guide.pdf>.

<sup>33</sup> *See id.*

<sup>34</sup> *See id.* at 16.

<sup>35</sup> Bennett Borden, Monica McCarroll, Mark Cordover & Sam Strickland, *Why Document Review is Broken*, WILLIAMS MULLEN (May 16, 2011), <http://www.williamsmullen.com/resources/detail.aspx?pub=664>.

review to specific custodians, limiting the data set by date range, and using search term filtering to separate the relevant documents from the rest of the data set.<sup>36</sup> Each of these criteria removes documents from the document set that humans will review. Of this list, search term filtering is probably the most complex because parties may have difficulty reaching an agreement on keywords during the meet and confer process.<sup>37</sup> Furthermore, the proper use of search term filtering is heavily dependent on technology and may require expertise in “statistics and linguistics.”<sup>38</sup> Search terms are generally developed by interviews with custodians and negotiations with the opposing party.<sup>39</sup> Search terms can effectively identify relevant documents in some cases, but it is difficult to balance precision and recall.<sup>40</sup> While search terms will limit the size of the data

---

<sup>36</sup> See SEDONA PRINCIPLES, *supra* note 29, at 38.

<sup>37</sup> See, e.g., *United States v. O’Keefe*, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (discussing how defendants “contended the search terms used by the government were insufficient”).

<sup>38</sup> *Id.*

<sup>39</sup> See FED. R. CIV. P. 26(f); see also SEDONA PRINCIPLES, *supra* note 29, at 21 (listing among the topics for the 26(f) conference, the “use of search terms and other methods of reducing the volume of electronically stored information to be preserved or produced”).

<sup>40</sup> See BRENT R. ROWE ET AL., ECONOMIC IMPACT ASSESSMENT OF NIST’S TEXT RETRIEVAL CONFERENCE (TREC) PROGRAM at § 2-4 (July 2010), *available at*

set, even after search term filtering, there is often a high percentage of irrelevant documents mixed in with the relevant documents that humans need to review.<sup>41</sup>

To reduce cost, litigators need the ability to analyze the collected data and identify the relevant documents that must be produced in order to check the documents for privilege, trade secrets, or categorize the documents in preparation for depositions or trial.<sup>42</sup> In addition, the producing party typically wants to know what they are producing before their opponent receives the production, since not all relevant documents have equal evidentiary value. There are many reasons a litigator may want humans to review the relevant documents before producing them to the opposing party. However, aside from confirming the absence of relevant documents, litigants gain little benefit from human review of

---

<http://trec.nist.gov> (explaining that precision refers to the percentage of relevant documents retrieved in a search while recall refers to the volume of irrelevant material that is also retrieved in the search).

<sup>41</sup> *Sedona Search Commentary*, *supra* note 6, at 199.

<sup>42</sup> *See id.* at 198.

irrelevant documents.<sup>43</sup> Historically, there has not been a reliable, industry-accepted technology or methodology to distinguish between relevant and irrelevant documents other than human review, meaning that humans necessarily needed to review a high percentage of documents that were not relevant.<sup>44</sup> In recent years, however, technology has improved.

#### B. The History of Technology-Assisted Review

Integrating technology with human review is not a new concept. Since the early days of reviewing electronic documents, attorneys commonly used technology to streamline and prioritize documents for human review and to assist in the quality control process.<sup>45</sup>

Early examples of integrating technology into human review include using a coding form to capture the reviewing attorney's work product, and using technology to create discreet batches of documents for humans to review.<sup>46</sup> Software providers developed indexing engines that

---

<sup>43</sup> *See id.* at 199.

<sup>44</sup> *Cf. id.* at 208.

<sup>45</sup> *Cf. id.* at 199.

<sup>46</sup> *See id.* at 209.

turned document text into searchable databases with metadata filters that empowered attorneys to organize documents by date range, custodian, and email thread or file type.<sup>47</sup> Attorneys could organize the documents in batches using date filters or search terms and prioritize batches for review.<sup>48</sup> Administrators gained the ability to set up workflows that allowed document reviewers to “check out” a batch, complete review, then “check in” the completed batch.<sup>49</sup>

Another type of technological advance used “checksum” or “hashing” algorithms to identify duplicate documents and remove the duplicate datasets (“de-dupe”), eliminating the need for attorneys to review identical documents, while retaining information about where those duplicate documents are located in the data set so these duplicates could be repopulated for production.<sup>50</sup> More recent technology identifies

---

<sup>47</sup> See *Sedona Search Commentary*, *supra* note 6, at 207-08.

<sup>48</sup> See *id.* at 200-01.

<sup>49</sup> See, e.g., Press Release, kCura, kCura Releases Relativity Six (May 3, 2010), available at <http://kcura.com/relativity/news/id/87/kcura-releases-e-discovery-review-platform-relativity-six>.

<sup>50</sup> See Craig Ball, *Meeting the Challenge: E-Mail in Civil Discovery*, in 5TH ANNUAL ADVANCED E-DISCOVERY INSTITUTE: THE DISCOVERY OF ESI COMES OF AGE 2008, 2008

not just exact duplicates, but “near-duplicates” that vary by a few words or sentences.<sup>51</sup>

What these tools have in common is their ability to assist or augment the human review process – meaning these tools assist the human process rather than remove unique documents from the set of documents a human would review. More recent technologies can limit the number of documents reviewed by humans by categorizing document sets and, under the right circumstances, culling out documents not likely to be relevant.<sup>52</sup> This technology permits the human review team to focus on the documents that are most likely to be relevant by limiting the documents that need to be reviewed or by categorizing the relevant documents for priority review.

---

WL 6654666, \*12 (Nov. 20, 2008); Ralph C. Losey, *Hash: The New Bates Stamp*, 12 J. TECH. L. & POL’Y 1, 12 (2007).

<sup>51</sup> *Sedona Search Commentary*, *supra* note 6, at 200.

<sup>52</sup> *See id.*

### C. Next Generation Technology-Assisted Review Tools

Earlier, we referenced some of the common criteria or tools used to limit the volume of documents that humans must review.<sup>53</sup> Technology allows us to filter data by custodian, date range and search terms.<sup>54</sup> While keyword searching (and for that matter, Boolean, fuzzy and concept searching) can prove effective, it has some limitations. For example, keyword searches are most effective when executed in iterations; however, the litigation process is not well suited to iterative keyword searching.<sup>55</sup> As a result the emergence of several newer technologies that do a better job at balancing precision and recall, and more reliably reduce the number of documents humans must review, is generating robust debate.<sup>56</sup>

---

<sup>53</sup> See *supra* Part II.B.

<sup>54</sup> See *supra* text accompanying note 47.

<sup>55</sup> John W. Woods, Lisa J. Sotto & Meghan A. Podolny, *Internal Investigation and eDiscovery*, in *EDISCOVERY FOR CORPORATE COUNSEL* §7:19 (Carole Basri & Mary Mack eds., 2011).

<sup>56</sup> See *Sedona Search Commentary*, *supra* note 6, at 194; see generally Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, XVII Rich. J.L. & Tech. 11 (2011), <http://jolt.richmond.edu/v17i3/article11.pdf>.

Technology-assisted review tools generally work by using a human to train a computer on the categories of documents the computer should identify as relevant.<sup>57</sup> The computer then quickly goes out and reviews the entire data set and categorizes documents as either relevant or irrelevant based on the training it received.<sup>58</sup> Humans then review the documents the computer identifies as relevant as well as a statistically significant sample of the documents the computer identified as not relevant to confirm that they are, in fact, not relevant.<sup>59</sup> Using statistical

---

<sup>57</sup> Maura Grossman and Gordon Cormack defined these types of technology-assisted review tools as follows:

*A technology-assisted review process involves the interplay of humans and computers to identify the documents in a collection that are responsive to a production request, or to identify those documents that should be withheld on the basis of privilege. A human examines and codes only those documents the computer identifies – a tiny fraction of the entire collection. Using the results of this human review, the computer codes the remaining documents in the collection for responsiveness (or privilege). A technology-assisted review process may involve, in whole or in part, the use of one or more approaches including, but not limited to, keyword search, Boolean search, conceptual search, clustering, machine learning, relevance ranking, and sampling.*

Grossman & Cormack, *supra* note 56.

<sup>58</sup> *Id.*

<sup>59</sup> The Sedona Conference, *The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, 10 SEDONA CONF. J. 299, 312 (2009).



models that long predate the existence of electronic discovery, we can quantify our confidence level and say with a specific degree of certainty that we have identified at least a specific percentage of the relevant documents.<sup>60</sup>

However, there is much more to using this type of technology-assisted review than pushing a button. As technology-assisted review tools developed, processes and safeguards for defensibly implementing these tools developed as well.<sup>61</sup> With older technologies, the process was more about efficiency than reliability, since humans ultimately reviewed each document in the data set, and technology impacted only the order in which the humans reviewed documents.<sup>62</sup> While some commentators dispute whether humans work more accurately than machines, there are fewer challenges to processes involving humans than to processes

---

<sup>60</sup> See Grossman & Cormack, *supra* note 56, at 44-46; *Sedona Search Commentary*, *supra* note 6, at 192.

<sup>61</sup> *Sedona Search Commentary*, *supra* note 6, at 199.

<sup>62</sup> See *id.* at 198-99.

involving machines – perhaps because of the prevailing belief that human review serves as the gold standard.<sup>63</sup>

With newer technologies that reduce the need for humans to look at every document, the several types of tools available must be used properly in order to achieve a reliable result. Even the best technology in the wrong hands is a recipe for disaster. Technology is only reliable when it is used in conjunction with the right process. Indeed, in the context of litigation, the process is just as important, and perhaps more important,

---

<sup>63</sup> *Id.* at 199 (“[T]here appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible – perhaps even perfect – and constitutes the gold standard by which all searches should be measured. Even assuming that the profession had the time and resources to continue to conduct manual review of massive sets of electronic data sets (which it does not), the relative efficacy of that approach versus utilizing newly developed automated methods of review remains very much open to debate.”). *But see generally* William Webber, *Re-examining the Effectiveness of Manual Review*, <http://www.umiacs.umd.edu/~oard/sire11/papers/webber.pdf> (last visited July 28, 2011) (revisiting the analysis in two well known articles, “Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review” and “Technology-assisted review in E-Discovery Can be More Effective And More Efficient than Exhaustive Manual Review,” which concluded, respectively, that manual review is at least as consistent as automated review and that manual review is superior to automated review). Webber argues that the previous studies, “while suggestive, are not conclusive” and calls for additional studies to answer the open question of whether an automated system can surpass or even achieve the reliability of a properly managed review team. *Id.* at 1.

than the technology. Some companies have taken these processes seriously enough to obtain patent protection for their technology.<sup>64</sup>

Historically, technology-assisted review augmented human review by allowing humans to review documents more efficiently.<sup>65</sup> The emerging generation of technology-assisted review tools is more analogous to search terms as it removes documents from the set of documents identified for human review.<sup>66</sup> However, using technology to limit the population of documents that will be reviewed by humans, either through the use of search terms or technology-assisted review, raises the question of reliability.

#### D. Is Technology-Assisted Review Reliable?

An attorney can assess the reliability of technology-assisted review tools the same way she assesses the reliability of search terms. Search terms are typically selected based on discussions with individual

---

<sup>64</sup> See, e.g., Press Release, Recommind, Inc., Recommind Patents Predictive Coding (June 8, 2011), available at [http://www.recommind.com/releases/20110608/recommind\\_patents\\_predictive\\_coding](http://www.recommind.com/releases/20110608/recommind_patents_predictive_coding).

<sup>65</sup> Cf. *Sedona Search Commentary*, supra note 6, at 193 (explaining the changes technology has caused to the discovery process).

<sup>66</sup> See *id.* at 201.

custodians regarding the terms likely to identify relevant documents and negotiations with the opposing party.<sup>67</sup> However, the only way to determine the actual precision and recall of the search terms is to review a statistical sample of the documents identified as relevant and not relevant to confirm the level of precision and recall.<sup>68</sup> Technology-assisted review tools require a similar statistical review.<sup>69</sup>

Sampling allows the producing party to review a subset of the corpus of documents the technology-assisted review tool identifies as not relevant and say with a statistical degree of certainty that the tool has located a statistical percentage of the relevant documents (depending on how many documents were reviewed and how many errors were

---

<sup>67</sup> The discussion of search terms is generally seen as a required topic at the Rule 26(f) conference. The authors strongly suggest that counsel likewise discuss plans to use technology-assisted review tools in lieu of human document review. *See supra* text accompanying note 38.

<sup>68</sup> *See* Grossman & Cormack, *supra* note 56, at 8 (explaining the commonly-used terms “recall” and “precision” of an information retrieval process as the “completeness” and “accuracy” of the search, respectively).

<sup>69</sup> *See, e.g., Application of Sampling to E-Discovery Search Result Evaluation*, E-DISCOVERY REFERENCE MODEL, app. 2, <http://www.edrm.net/resources/guides/edrm-search-guide/appendix-2> (last visited Feb. 4, 2012).

identified).<sup>70</sup> Thus, whether the use of a technology-assisted review tool was defensible will come down to whether the party that used the technology can demonstrate that they followed the process and produced reliable results.

III. TECHNOLOGY-ASSISTED REVIEW IS DEFENSIBLE AND CONSISTENT WITH FUNDAMENTAL PURPOSES OF THE FEDERAL RULES OF CIVIL PROCEDURE

Evaluating the reasonableness of technology-assisted review in any given case requires the evaluation of two considerations: defensibility and proportionality.<sup>71</sup> The process employed must, in the first instance, be defensible, meaning that the proponent can “demonstrate to opposing parties, courts, and government agencies, that its chosen method and tool accurately captured a reasonably sufficient number of the relevant, nonprivileged ESI in existence, and that the remaining unreviewed and

---

<sup>70</sup> *See id.* (“The estimate of the proportion of responsive documents from a random sample can be stated to be within a specified number of standard deviations from the sample’s proportion with a specific confidence level.”).

<sup>71</sup> *See* Grossman & Cormack, *supra* note 56, at ¶ 5; *see also* The Sedona Conference, *The Sedona Conference Commentary on Proportionality in Electronic Discovery*, 11 SEDONA CONF. J. 289, 294 (2010).

unproduced ESI is irrelevant.”<sup>72</sup> In addition, the proponent of technology-assisted review must demonstrate that its process satisfies the requirements of Rules 1 and 26 of the Federal Rules of Civil Procedure: that “the burden or expense of...discovery [does not] outweigh[] its likely benefit”<sup>73</sup> and helps to secure “the just, speedy, and inexpensive determination of [the] action[.]”<sup>74</sup>

#### A. Technology-Assisted Review is Defensible

To date, no court has addressed the defensibility of using technology-assisted review to exclude from review and production documents unlikely to contain relevant information.<sup>75</sup> Nevertheless,

---

<sup>72</sup> The Sedona Conference, *The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, 10 SEDONA CONF. J. 299, 320 (2009).

<sup>73</sup> FED. R. CIV. P. 1.

<sup>74</sup> FED. R. CIV. P. 26(b)(2)(C)(iii).

<sup>75</sup> See, e.g., Peck, *supra* note 16 (“[N]o reported case (federal or state) has ruled on the use of computer-assisted coding. While anecdotally it appears that some lawyers are using predictive coding technology, it also appears that many lawyers (and their clients) are waiting for a judicial decision approving of computer-assisted review.”) If faced with a challenge to the use of technology-assisted review, Judge Peck would first consider “what was done and why that produced defensible results,” focusing on “whether [the process] produced responsive documents with reasonably high recall and high precision.” See *id.* See also, Peck, *supra* note 16.

technology-assisted review is consistent with existing jurisprudence on the defensibility of using technology to facilitate the discovery of ESI.

In *Victor Stanley, Inc. v. Creative Pipe, Inc.*, Chief Magistrate Judge Paul Grimm analyzed a discovery dispute involving the inadvertent production of 165 purportedly privileged documents.<sup>76</sup> Judge Grimm ruled the attorneys had waived privilege and failed to prove they had undertaken a reasonable privilege review process.<sup>77</sup> In assessing the adequacy of the defendants' privilege review process, Judge Grimm noted that "it is universally acknowledged that keyword searches are useful tools for search and retrieval of ESI."<sup>78</sup> He further noted the danger of using "an unreliable or inadequate keyword search," and emphasized the importance of sampling to "test the reliability of the keyword search."<sup>79</sup>

---

<sup>76</sup> 250 F.R.D. 251, 253 (D. Md. 2008).

<sup>77</sup> *See id.* at 257-59, 262.

<sup>78</sup> *Id.* at 256.

<sup>79</sup> *See id.* at 257 ("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.").

In that case, the defendants failed to provide any evidence to support the reliability of their keyword search for privileged documents:

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 [one of the defendants] 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.<sup>80</sup>

Similarly, in *Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, a case involving alleged violations of the Americans with Disabilities Act, Magistrate Judge John Facciola was asked to resolve a dispute about whether the defendant should search backup tapes for information deleted from its computer systems during the course of the litigation.<sup>81</sup> The defendant objected on the basis that the process would be

---

<sup>80</sup> *Id.* at 259-60.

<sup>81</sup> 242 F.R.D. 139, 145-46 (D.D.C. 2007).



unduly burdensome and expensive.<sup>82</sup> Because potentially relevant information had been deleted after the duty to preserve had arisen and would only exist, if at all, on backup tapes, Judge Facciola ordered the defendant to search the tapes.<sup>83</sup> Further, he ordered the parties to confer on the process by which the backup tapes would be searched, noting that “recent scholarship...argues that concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results.”<sup>84</sup> Thus, courts have recognized that the proper use of technology to improve the quality and efficiency of document review is defensible so long as the proponent of the methodology can

---

<sup>82</sup> *Id.* at 147-48; *see also* FED. R. CIV. P. 26(b)(2)(B) (“A party need not provide discovery of electronically stored information from sources that the party identifies as not reasonably accessible because of undue burden or cost.”).

<sup>83</sup> *See Disability Rights Council*, 242 F.R.D. at 147-48.

<sup>84</sup> *See id.* at 148; *see also* A.N.S.W.E.R. Coal. v. Salazar, 258 F.R.D. 36, 38 (D.D.C. 2009) (ordering parties to confer on “a methodology for [keyword] searches [and] . . . a list of search directives that are likely to result in [relevant] documents”); *Am. Family Mut. Ins. Co. v. Gustafson*, No. 08-cv-02772-MSK-MJW, 2009 WL 641297, at \*3 (D. Colo. Mar. 10, 2009) (ordering parties to “meet, confer, and agree upon the search terms that will be used” to search imaged hard drive).

explain “the [methodology] used; [and] the rationale for [its] selection . . .

.”<sup>85</sup>

B. Technology-Assisted Review Furthers the Goal that the  
Burden and Expense of Discovery be Proportional  
to the Needs of the Case

Discovery “is defined in the first instance by relevance to the claims and defenses in a case.”<sup>86</sup> Though “the bounds of permissible discovery in a civil action are generally regarded as expansive . . . they are not without limits.”<sup>87</sup> In addition, a “party need not provide discovery of electronically stored information from sources that the party identifies as not reasonably accessible because of undue burden or cost.”<sup>88</sup> Indeed, a court *must* limit discovery “if it determines that the burden or expense of the proposed discovery outweighs its likely benefit, considering the needs of the case, the amount in controversy, the parties’ resources, the

---

<sup>85</sup> See *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 259-60 (D. Md. 2008).

<sup>86</sup> *Wood v. Capital One Serv., LLC.*, No. 5:09-CV-1445 (NPM/DEP), 2011 WL 2154279, at \*5 (N.D.N.Y. Apr. 15, 2011).

<sup>87</sup> *Id.* at \*3; see also *Averett v. Honda of Am. Mfg., Inc.*, No. 2:07-cv-1167, 2009 WL 799638, at \*2 (S.D. Ohio Mar. 24, 2009) (“recent revisions [to the Federal Rules of Civil Procedure] communicate the message that discovery is not unlimited”).

<sup>88</sup> FED. R. CIV. P. 26(b)(2)(B).

importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.”<sup>89</sup> Similarly, Rule 26(c) allows a court to protect a party against “undue burden or expense.”<sup>90</sup>

The plaintiff in *Wood v. Capital One Services, LLC* claimed the defendants violated the Federal Debt Collection Practices Act.<sup>91</sup> Prior to filing the motions discussed in the opinion, “the parties engaged in a considerable amount of discovery,” including numerous interrogatories and document requests directed to Capital One Services.<sup>92</sup> In response to the interrogatories, the defendant produced 1,500 pages of documents, and proffered a Rule 30(b)(6) witness for two days of deposition on several topics, including the methods the company had used to answer the

---

<sup>89</sup> FED. R. CIV. P. 26(b)(2)(C)(iii) (emphasis added); see *Dilley v. Metro. Life Ins. Co.*, 256 F.R.D. 643, 644 (N.D. Cal. 2009) (“The court must limit discovery if it determines that ‘the burden or expense of the proposed discovery outweighs its likely benefit,’ considering certain factors including ‘the importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.’”) (quoting FED. R. CIV. P. 26(b)(2)(C)(iii)); *Averett*, 2009 WL 799638, at \*2 (“the court always has a duty to limit discovery under Rule 26(b)(2)(C)(i)-(iii)”).

<sup>90</sup> See FED. R. CIV. P. 26(c).

<sup>91</sup> See *Wood*, 2011 WL 2154279, at \*1.

<sup>92</sup> See *id.* at \*6.

plaintiff's discovery request.<sup>93</sup> In his motion to compel, the plaintiff chose specific search terms and asked that Capital One Services use the terms to search the e-mail accounts of forty-one employees.<sup>94</sup> Capital One Services established that "the likely volume to be generated by the requested searches, after elimination of duplicates, is as high as 1,753,537 documents, costing in excess of \$5,000,000 to process, review, and produce."<sup>95</sup>

The court denied the plaintiff's motion and granted Capital One Services's motion for protective order, relying on Rule 26(b)(2)(C)(iii), which, the court stated, "serves to protect a party against having to produce voluminous documents of questionable relevance."<sup>96</sup> The court found that the plaintiff had failed to "shed significant light on the potential relevance of the documents sought," while the defendants had "clearly identified an inordinate burden associated with responding to the

---

<sup>93</sup> *See id.*

<sup>94</sup> *See id.* at \*7.

<sup>95</sup> *See id.* at \*8.

<sup>96</sup> *See Wood v. Capital One Serv., LLC.*, No. 5:09-CV-1445 (NPM/DEP), 2011 WL 2154279, at \*3 (N.D.N.Y. Apr. 15, 2011).

request.”<sup>97</sup> As a result, the “rule of proportionality” dictated that the plaintiff’s motion be denied “without prejudice to his right to renew the motion to compel in the event he is willing to underwrite the expense associated with any such search.”<sup>98</sup>

A similar result was obtained in *Daugherty v. Murphy*.<sup>99</sup> In that class action case alleging violations of due process and federal and state law in the handling of Medicaid claims, the parties brought before the court a dispute over the defendants’ production of extracts from a certain computer system.<sup>100</sup> The defendants moved for a protective order and the plaintiffs filed a motion to compel, each asking the court to order production of the data extracts outlined in their competing proposals.<sup>101</sup>

The court first outlined the law governing the opposing motions:

While the scope of discovery is broad under Rule 26, that rule confers broad powers on the court to regulate or deny discovery even though the materials sought are otherwise

---

<sup>97</sup> *See id.* at \*8-9.

<sup>98</sup> *See id.* at \*7, \*9.

<sup>99</sup> No. 1:06-cv-0878-SEB-DML, 2010 WL 4877720, at \*1 (S.D. Ind. Nov. 23, 2010).

<sup>100</sup> *Id.* at \*1-3.

<sup>101</sup> *Id.* at \*3-4.

within the scope of Rule 26(b)(1). Rule 26(b) provides that the scope of discovery may be ‘limited by court order,’ and Rule 26(b)(2)(C) requires the court to limit discovery if the court determines that the burden or expense of the discovery on one party outweighs its likely benefit to the other party, after considering ‘the needs of the case, the amount in controversy, the parties’ resources, the importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.’<sup>102</sup>

According to the defendants, their proposed data extract would cost \$36,000, of which the defendants had already spent \$16,000.<sup>103</sup> In contrast, plaintiffs’ proposed data extract would cost nearly \$100,000.<sup>104</sup> Although the plaintiffs argued that the cost was exaggerated, the court disagreed:

When the court compares the heavy time and expense to create the data extracts that the plaintiffs originally proposed (and assuming that their new proposal will request a similar number of extracts) with the benefits of that discovery and its importance to the issues to be resolved in this case, the plaintiffs come up short. The plaintiffs have not provided a clear explanation of how the data from [the defendants’] extracts is insufficient to allow the plaintiffs to present evidence of the proper scope of

---

<sup>102</sup> *Id.* at \*4; FED. R. CIV. P. 26(b)(2)(C)(iii).

<sup>103</sup> *Daugherty v. Murphy*, No. 1:06-cv-0878-SEB-DML, 2010 WL 4877720, at \*5 (S.D. Ind. Nov. 23, 2010).

<sup>104</sup> *Id.*

Class 1 and/or the proper injunctive relief for Class 1. Simply asserting that their expert would like to have it is not enough.<sup>105</sup>

Further, the court denied the plaintiffs' request to allow their expert to design new data extracts because the plaintiffs failed to "convince the court that the burdens and benefits of data extracts should be measured dramatically differently."<sup>106</sup>

#### IV. DEFENDANT'S TECHNOLOGY-ASSISTED REVIEW WAS DEFENSIBLE AND PROPORTIONAL TO THE NEEDS OF THE LITIGATION

Defendant deployed technology-assisted review in a reasonable manner in this case because: Defendant described the process in sufficient detail; Defendant balanced privilege against pressing time deadlines and the requirement to cooperate; and Defendant has implemented sufficient safeguards and quality control mechanisms to meet the standards set forth in the Rules and the case law interpreting the Rules.<sup>107</sup> Defendant should

---

<sup>105</sup> *Id.* at \*7.

<sup>106</sup> *Id.* at \*8.

<sup>107</sup> As discussed at the start of this article, a detailed recitation of the necessary safeguards, Judge Peck's "careful thought, quality control, testing and cooperation" is beyond the scope of this article. *William A. Gross Constr. Assocs. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134, 134 (S.D.N.Y. 2009); *see supra* Part I.

not be sanctioned for using technology-assisted review and avoiding the human review of documents, less than 3% of which may be relevant to this case.<sup>108</sup>

Technologies like de-duplication and keyword searching have become acceptable tools to limit the volume of documents reviewed by humans in the discovery process.<sup>109</sup> Technology-assisted review is yet another reasonable and defensible method of reducing the volume of documents designated for human review prior to production.<sup>110</sup>

Defendant's use of technology-assisted review in this case is reasonable because Defendant has balanced the technology's limitations with reasonable safeguards. Defendant reviewed a statistically significant sample of the 800,000 documents that the technology identified as non-responsive after it was trained by human reviewers. That sampling has yielded a 95% confidence level that less than 3% of the documents are

---

<sup>108</sup> See *supra* Part I.

<sup>109</sup> See John Markoff, *Armies of Expensive Lawyers Replaced by Cheaper Software*, N.Y. TIMES, Mar. 4, 2011, <http://www.nytimes.com/2011/03/05/science/05legal.html?pagewanted=all>.

<sup>110</sup> See *id.*



relevant.<sup>111</sup> The cost of paying attorneys to review 800,000 additional documents to find 24,000 potentially relevant documents is overly burdensome and disproportionate under the Rules.<sup>112</sup>

Plaintiff inaccurately describes technology-assisted review as a technology and a process that *replaces* contract attorneys and non-attorney review professionals who currently perform document review.<sup>113</sup> *Replacement* implies a “silver bullet” solution that over-simplifies the discovery process and ignores the role that attorney reviewers have played in this case.<sup>114</sup> Plaintiff ignores a critical fact: Defendant has not *replaced* human lawyers. Defendant has instead incorporated technology-assisted review into a process that has remained under the control of counsel. Human reviewers designed and executed a document review plan that leveraged technology to meet the requirements of defensibility and proportionality. Humans reviewed the documents that gave the

---

<sup>111</sup> See *supra* Part I.

<sup>112</sup> See FED. R. CIV. P. 26(b)(2)(C)(iii).

<sup>113</sup> See Markoff, *supra* note 109.

<sup>114</sup> See *Benefits and Risks of Predictive Coding*, EXECUTIVE COUNS. INST., <http://www.executivecounselinstitute.com/e-discovery/benefits-and-risks-of-predictive-coding> (last visited Feb. 2, 2012).

technology the input to predictively code the data set, the documents that the technology-assisted review tool identified as responsive or privileged, and a statistically significant sample of the documents identified as non-responsive as well. Finally, information from the documents, the information counsel is using to develop trial themes, has been of necessity transferred to the lawyers who are preparing to try the case. The better question is whether attorneys are using technology-assisted review in a defensible process, leveraging the right expertise and with appropriate safeguards to improve certain phases of the discovery process, not whether technology is *replacing* lawyers. The reasonableness of deploying technology in electronic discovery with appropriate safeguards has already been answered affirmatively for technologies like de-duplication and keyword searching.<sup>115</sup> As courts have ruled in prior cases like *Victor Stanley, O'Keefe*, and *In re Seroquel*, using technology to assist attorneys in the discovery process is defensible as long as it is implemented with sufficient safeguards and documentation.<sup>116</sup>

---

<sup>115</sup> *Id.*

<sup>116</sup> *See, e.g.*, FED. R. EVID. 502 Committee Note (“Depending on the circumstances, a party that uses advanced analytical software applications and linguistic tools in screening

Defendant has met its burden of establishing sufficient safeguards for using technology-assisted review and comparing such safeguards it to its use of technology in this case.

#### V. CONCLUSION OF HYPOTHETICAL BRIEF

The Court should deny Plaintiff's motion to compel, and grant Defendant's motion for protective order because the technology-assisted was reasonable, defensible, and the burden and expense of Plaintiff's requested relief would far outweigh the likely benefit.

### III. ARTICLE CONCLUSION

[7] Lawyers do not eliminate risk; we manage risk. Increasingly, attorneys must develop a better understanding of the technology our clients use to generate potential evidence, and the technology available to sort through voluminous data to find necessary information. Trying cases still means developing trial themes and presenting evidence to support the elements of claims and defenses. That remains a human process. The process we use to identify, preserve, review and produce information in discovery is still evaluated based on a reasonableness standard. Reasonableness is still evaluated based on the "reasonable person" standard, not the "reasonable computer."

[8] We cannot say and do not attempt to forecast how any given judge would rule on this brief. What we can say is that given what we know

---

for privilege and work product may be found to have taken 'reasonable steps' to prevent inadvertent disclosure."); *United States v. O'Keefe*, 537 F. Supp. 2d. 14, 18, 23-24 (D.D.C. 2008); *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 259 (D. Md. 2008); *In re Seroquel*, 244 F.R.D. 650, 663 (M.D. Fla. 2007).

about the current state of the federal rules and case law, taken in conjunction with statements by thought leaders and leading members of the judiciary, technology-assisted review, when implemented with the right expertise and sufficient safeguards, can be reliable when used in conjunction with the right process.

[9] We can say that the analysis depends on the facts of each case. Those facts flow directly from a new reality lawyers must accept: in addition to a thorough understanding of a developing body of case law, attorneys can no longer hide their heads in the sand and ignore technology. They must actively pursue an understanding of the technology their clients use to run their business, and the technology that preserves and re-formats data for use in court or staff their litigation team with lawyers that do understand it. Technology and electronic discovery will always present challenges, but lawyers must respond to the challenge by understanding, or finding someone who understands, how tools like technology-assisted review work. Heeding the “wake-up call” and developing expertise in litigation technologies and the processes to implement them will allow lawyers to change the way we think of electronic discovery – electronic discovery should be a solution, not a problem.