

MANDATING USE OF PREDICTIVE CODING IN ELECTRONIC DISCOVERY:  
AN ILL-ADVISED JUDICIAL INTRUSION

By

Tonia Hap Murphy\*

“But lo! Men have become the tools of their tools.”  
--Henry David Thoreau<sup>1</sup>

## I. Introduction

In twenty-first century civil litigation, discovery focuses on electronically stored information (“ESI”).<sup>2</sup> Lawsuits may be won or lost because of incriminating electronic documents.<sup>3</sup> While offices are moving toward less use of paper, the amount of ESI potentially discoverable in litigation has proliferated.<sup>4</sup> Litigants in some cases may face the task of searching billions of electronic records for those documents that are responsive to discovery requests.<sup>5</sup> That makes discovery a burdensome and expensive proposition.<sup>6</sup>

While production of documents may evoke images of young associates in dusty storerooms looking at thousands of individual paper documents stored in boxes or filing cabinets, discovery has changed with the burgeoning of ESI. Computers now do much of the work, culling from that vast body of ESI a smaller set of documents that may be responsive and relevant. For at least the past decade, most electronic discovery has been via keyword searching<sup>7</sup> -- a great improvement over old-fashioned manual review, but presenting its own problems. A newer generation of search technology is “predictive coding.” Despite advocates’ claims of improved accuracy and lower costs, attorneys have been slow to transition to predictive coding methods. Moreover, there has been a lack of precedent indicating that judges regard predictive coding as reliable and acceptable under the Federal Rules of Civil Procedure. Adjustment to advances in technology often come slowly, but good reasons may exist to proceed cautiously here, especially in lawsuits where large amounts of money or even the existence of a corporation may be at stake.

Two cases, *Da Silva Moore v. Publicis Groupe*<sup>8</sup> and *Kleen Products, LLC v. Packaging Corporation of America*,<sup>9</sup> suddenly have brought predictive coding to the forefront in the spring of 2012. In these cases, federal magistrate judges have signaled approval of predictive coding and may be ready to mandate its use, rather than keyword searching methods, over parties’ objections. A state court in Virginia also ruled in April 2012, permitting the use of predictive coding over a party’s objection.<sup>10</sup> These cases have attracted considerable attention, especially in law and technology blogs.<sup>11</sup> This paper seeks to provide a detached, deeper, scholarly analysis, to provide guidance as other courts may grapple with this novel issue.

After examining in greater detail these two types of technology, the facts and rulings of *Da Silva Moore* and *Kleen Products*, and the fundamental roles of the parties and the court in discovery, this paper will argue that judicial mandates of particular technologies for electronic discovery would be misguided because the parties may have legitimate, good faith concerns that lead them to prefer keyword searching rather than predictive coding and because judges do not have sufficient reasons to depart from the traditional judicial role to intervene in such decisions.

## II. Technology for Electronic Discovery: Keyword Searching to Predictive Coding

Keyword searches used in discovery are akin to the keyword searches used to find relevant documents on LEXIS or Westlaw. In both contexts, Boolean logic is used to refine searches by the use of connectors, such as “and,” “but not,” and “within same sentence.” Keyword searches in electronic discovery, however, are typically much more complex. In addition to Boolean logic, they may incorporate other advanced information retrieval methods.<sup>12</sup> The parties work together to define appropriate search protocols and may also limit the universe of ESI to be searched.<sup>13</sup> Largely to control costs, document review may be outsourced to attorneys on the other side of the globe.<sup>14</sup> Figure 1 depicts the discovery process when keyword searching is employed.<sup>15</sup>

Keyword searching is now the most widely used method of combing ESI for relevant documents,<sup>16</sup> but the method is not perfect. First, parties often disagree about appropriate search terms, which may cause the court to issue sanctions for lack of good faith cooperation in the process<sup>17</sup> or even court involvement in crafting search terms.<sup>18</sup> The major difficulty, however, is that parties may craft search terms that are over- or under-inclusive, either returning large amounts of irrelevant documents or failing to capture relevant ones.<sup>19</sup> Appropriately inclusive results are more likely when technical experts are involved in developing search terms.<sup>20</sup> Courts may require the testimony of experts on how the search terms were crafted.<sup>21</sup> And of course, the cost of the process is a concern.

---

\* Associate Teaching Professor, Mendoza College of Business, University of Notre Dame.

**Figure 1. Keyword Searching Process.**



The next generation of technology for electronic discovery is predictive coding.<sup>22</sup> A Rand Institute for Civil Justice monograph describes the approach:

Predictive coding . . . is a process by which the computer does the heavy lifting in deciding whether documents are relevant, responsive or privileged. . . . [The computer] automatically assign[s] a rating (or *proximity score*) to each document to reflect how close it is to the concepts and terms found in examples of documents attorneys have already determined to be relevant, responsive, or privileged. This assignment becomes increasingly accurate as the software continues to learn from human reviewers about what is, and what is not, of interest. This score and the self-learning function are the two key characteristics that set predictive coding apart from less robust analytical techniques.<sup>23</sup>

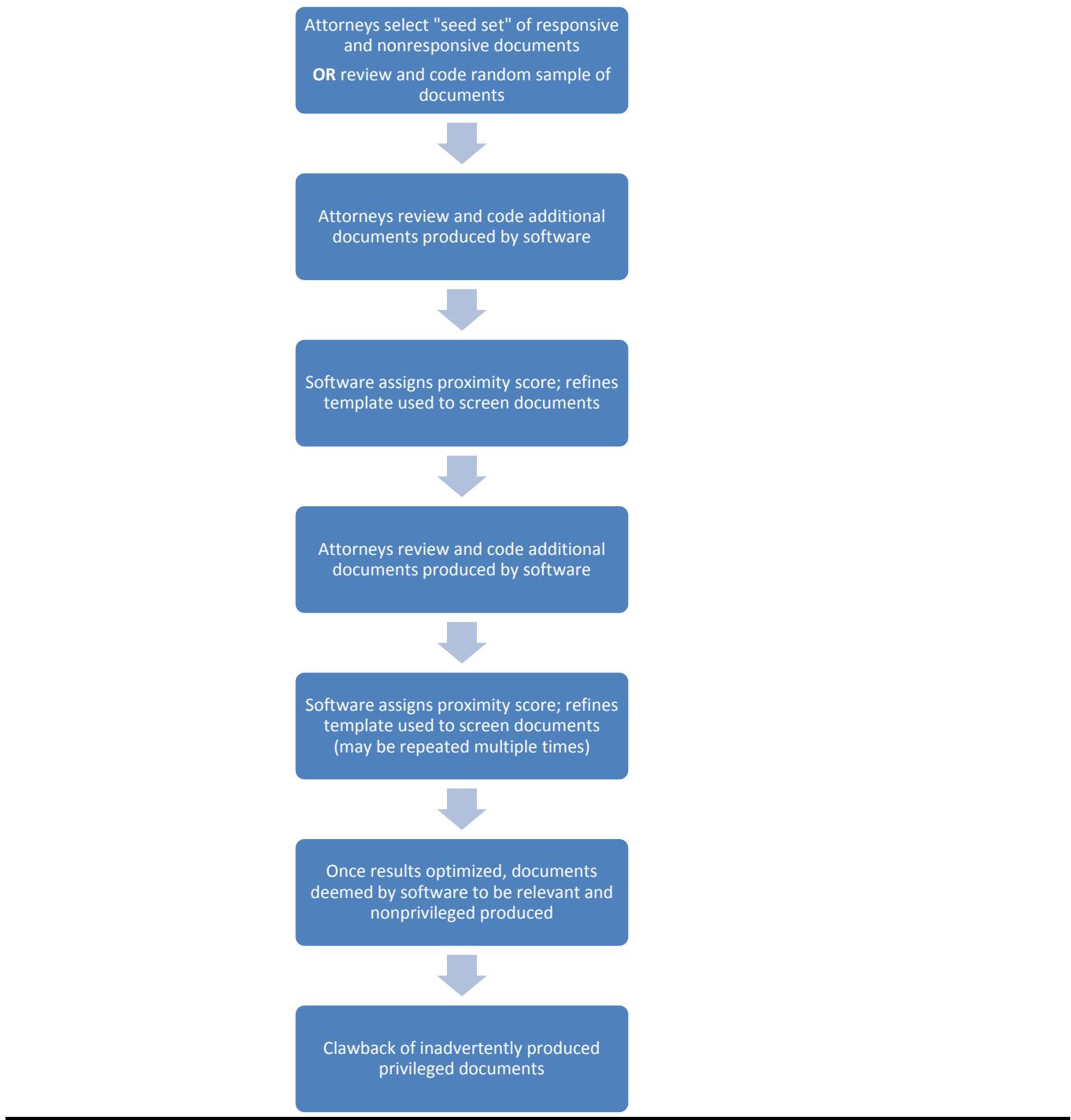
Attorneys, typically senior attorneys, work to train or calibrate the predictive coding software. The process begins either with attorneys selecting a “seed set” of responsive and non-responsive documents, or reviewing and coding a random sample of documents. These initial documents are then analyzed by the predictive coding software. The software begins to make judgments on probable relevance of other documents. The attorneys review further samples produced by the software, again applying their own judgment as to relevance, responsiveness, and privilege. The process continues until the attorneys are satisfied that the software is properly calibrated. At that point, the results are said to be “optimized”: “disagreement between the software’s decisions and those of the human reviewers should be kept to a minimum.”<sup>24</sup> Figure 2 depicts the discovery process when predictive coding is employed.<sup>25</sup>

Cost savings through predictive coding would be expected primarily by avoiding or at least greatly limiting final, attorney review of the computer-generated set of responsive documents.<sup>26</sup> Empirical data proving cost savings is scarce.<sup>27</sup> In assessing savings from use of predictive coding, one must account for vendor charges or licensing fees that may offset savings obtained due to fewer attorney hours.<sup>28</sup>

Exact methods for predictive coding are proprietary and vary from one provider to another. At least three companies have patents on technology or methods.<sup>29</sup> Other patent applications are pending.<sup>30</sup> Through published patent documents, details about such companies’ particular methods are publicly available. For other companies, general information about methodologies is available, but precise details on techniques are protected as trade secrets.<sup>31</sup> One would expect that effectiveness may vary from one provider to another.

Enthusiasm for predictive coding as a method for conducting electronic discovery runs high in some quarters,<sup>32</sup> but there are reasons to be cautious. First, there is a dearth of empirical evidence as to its effectiveness.<sup>33</sup> Limits of keyword searching may be overstated; in fact, there is overlap in the keyword searching and predictive coding approaches.<sup>34</sup> Additionally, as

**Figure 2. Predictive Coding Process.**



prominent information retrieval scholars have noted, “audacious claims [are] now being made in the E-discovery community regarding the efficacy of certain techniques.”<sup>35</sup> They cite “marketing claims that seem well beyond what we are presently able to measure, and perhaps even somewhat beyond what is actually possible.”<sup>36</sup>

Indeed, while predictive coding technology may have great promise, it has not yet been widely used in electronic discovery.<sup>37</sup> Observers ascribe slow adoption of predictive coding to several factors, including lack of adequate technical understanding by lawyers,<sup>38</sup> lack of transparency of the process,<sup>39</sup> and concern about accuracy of results.<sup>40</sup> Attorneys may also have concern about “the uncertainty of judicial acceptance”<sup>41</sup> and whether privilege protection may be lost for documents produced through the predictive coding process.<sup>42</sup>

### **III. Predictive Coding in the Courts: Moves to Mandate Use of that Technology**

Even in the early months of 2012, there were serious questions as to whether predictive coding would gain judicial acceptance. Could a party wishing to use predictive coding to cull responsive documents convince the court that the method will be sufficiently accurate in finding responsive documents? The Sedona Conference research institution opined that “if human review or even keyword searching is the benchmark for accuracy and reliability, it arguably should not be difficult to compare the new [concept searching] technology favorably . . . . The discovery standard is, after all, reasonableness, not perfection.”<sup>43</sup> But the idea was untested in the courts.

On February 24, 2012, however, Magistrate Judge Andrew Peck of the Southern District of New York ruled in *Da Silva Moore v. Publicis Groupe* that predictive coding “is an acceptable way to search for relevant ESI in appropriate cases.”<sup>44</sup> *Da Silva Moore* involves five female plaintiffs suing one of the world’s largest advertising firms and its public relations subsidiary for gender and pregnancy discrimination and various other claims.<sup>45</sup> The plaintiffs appealed Magistrate Judge Peck’s February 24 ruling to District Judge Andrew L. Carter, Jr., arguing among other things that the predictive coding method approved by Judge Peck was unreliable and unacceptable under the Federal Rules of Civil Procedure.<sup>46</sup> They also requested recusal of Judge Peck due to alleged appearance of partiality.<sup>47</sup> Noting a “highly deferential standard of review,” Judge Carter affirmed Judge Peck’s discovery ruling.<sup>48</sup> Judge Carter noted that while “the Plaintiffs and Judge Peck disagree about the scope of Plaintiffs’ acquiescence concerning the use of the [predictive coding] method,”<sup>49</sup> the plaintiffs had consented to its use, and “[a]t this stage, there is insufficient evidence to conclude that the use of the predictive coding software will deny Plaintiffs access to liberal discovery.”<sup>50</sup>

The other federal case of current interest is *Kleen Products, LLC v. Packaging Corporation of America*.<sup>51</sup> The case, pending in the Northern District of Illinois, is a class action antitrust claim, alleging price fixing in the containerboard industry by seven separate defendants.<sup>52</sup> District Judge Milton I. Shadur referred discovery to Magistrate Judge Nan Nolan. In their Joint Status Conference Report of May 17, 2012, the parties state that the plaintiffs have produced 25,000 pages of documents to date, and the defendants have produced more than 3,000,000 pages of documents, including both hard copies and ESI. Both parties expect to produce further documents.<sup>53</sup> Among the discovery issues is whether the keyword searching technology used by the defendants satisfies the Federal Rules of Civil Procedure. As stated in the defendant’s brief:

Plaintiffs have not meaningfully challenged the specific ESI search methodology that Defendants implemented here. Rather, Plaintiffs challenge the use of search terms *per se* in this litigation. Plaintiffs argue that Defendants, notwithstanding their very substantial investment of time and resources, must abandon what they have done and instead use the methodology Plaintiffs prefer.<sup>54</sup>

The plaintiffs in *Kleen Products* call the defendants’ search methodology “an antiquated Boolean keyword search” and request the judge to order the defendants to employ predictive coding.<sup>55</sup> After numerous days of evidentiary hearings and numerous continuances, as of May 29, 2012, there has been no ruling by Judge Nolan on the predictive coding issue.

Regardless of the ultimate rulings in these pending cases, *Da Silva Moore* and *Kleen Products* show this to be an important, new issue, likely to arise in other courts<sup>56</sup> and deserving of thorough, unbiased, scholarly attention, which this paper attempts to provide. A starting point for analysis is a look at the traditional roles of the parties and the court in discovery.

### **IV. Roles of the Parties and the Court in Discovery under the Federal Rules of Civil Procedure**

Discovery is designed to be a cooperative process, run by the parties, as represented by counsel, largely without interference by the court.<sup>57</sup> The American system of discovery is distinctive:

Party-initiated discovery is in stark contrast to civil law countries where (although changing) it is the civilian law judge who decides what evidence is needed and proceeds to request documents and interrogate witnesses in person, summarizing the testimony in writing. In the United States, it is the lawyers who conduct pretrial discovery, albeit supervised in a general way by judges when there is active case management.<sup>58</sup>

Two influential sources on “best practices” for electronic discovery, the Seventh Circuit Electronic Discovery Pilot Program and the Sedona Conference, characterize discovery as a cooperative, party-driven process. Principle 1.01 of the Seventh Circuit Pilot Program’s *General Principles* directs parties to work toward “the early resolution of disputes regarding the discovery of electronically stored information (‘ESI’) without Court intervention,” and Principle 1.02 reminds counsel that “An attorney’s zealous representation of a client is not compromised by conducting discovery in a cooperative manner.”<sup>59</sup> The Sedona Conference’s *Resources for the Judiciary* emphasize that “the Resources do not mean to imply that judges should be routinely making discovery decisions for the parties,” noting further that “Discovery is designed to be, and remains, party-driven. Active case management provides a strong framework in which the parties should develop and execute their own cooperative discovery plans.”<sup>60</sup>

Party-directed discovery, a hallmark of the American system, is thought to yield several benefits. First, the court’s limited involvement in the process conserves judicial resources. More importantly, parties who control aspects of the litigation

process are more likely to accept the court’s ultimate decision.<sup>61</sup> Party-directed discovery evidences respect for individual rights and opinions.<sup>62</sup> Additionally, it fortifies society’s confidence in the fairness of the adversary system:

When litigants direct the proceedings, there is little opportunity for the judge to pursue her own agenda or to act on her own biases. Because the judge seldom takes the lead in conducting the proceedings, she is unlikely to appear to be partisan or to become embroiled in the contest. Her detachment preserves the appearance of fairness as well as fairness itself.<sup>63</sup>

The discovery process commences under the Federal Rules of Civil Procedure when the parties “meet and confer” at a planning conference under Rule 26(f). At this conference, the parties must work together in good faith to develop a proposed discovery plan.<sup>64</sup> The plan shall address subjects for discovery and the timing of discovery.<sup>65</sup> The plan shall also address “issues about claims of privilege,” including whether the parties “agree on a procedure to assert these claims after production.”<sup>66</sup>

The parties submit the discovery plan to the court<sup>67</sup> for consideration and approval at a pre-trial conference. Rule 16 governs “Pre-Trial Conferences; Scheduling; [and] Management.” A guide for judges published by the Federal Judicial Center explains its purpose: “The Rule 16 conference and order afford the court the opportunity, early in the case, to discuss and memorialize the agreements or shared understandings that the parties reach in their ‘meet and confer’ session, and to resolve disputes that may have arisen.”<sup>68</sup> The judge issues an appropriate order, setting deadlines for such activities as pre-trial motions and discovery, memorializing the court’s rulings on such things as the extent of discovery, and memorializing any agreement of the parties on privilege protection.<sup>69</sup> Rule 16 specifically notes that the order may “provide for . . . discovery of electronically stored information.”<sup>70</sup>

Appropriate pre-trial functions for the court include “expediting disposition of the action,”<sup>71</sup> “establishing early and continuing control so that the case will not be protracted because of lack of management,”<sup>72</sup> and “discouraging wasteful pretrial activities.”<sup>73</sup> These functions align with the overarching goal of the Federal Rules of Civil Procedure, set out in Rule 1: “[These Rules] should be construed and administered to secure the just, speedy, and inexpensive determination of every actions and proceeding.”<sup>74</sup> Courts generally push parties to come to agreement between themselves on discovery issues.<sup>75</sup> Consistent with the idea that the court is there to supervise and not to take on primary responsibility for discovery, the court is advised merely to grasp “the relevant technology at a level that allows effective communication with attorneys, parties, and experts.”<sup>76</sup>

The role of the court is to supervise the discovery process and to intervene when it is abused by the parties.<sup>77</sup> Abuse occurs when parties violate the letter or spirit of the Rules through gamesmanship,<sup>78</sup> courtesy,<sup>79</sup> or inability or unwillingness to comply in an orderly fashion.<sup>80</sup> There is a general concern for fairness of the process.<sup>81</sup> Courts seek to balance the need to discover relevant, important information against the burdens of electronic discovery. Courts may be inclined to act especially when a party with fewer economic resources may be disadvantaged by the process.<sup>82</sup> Courts have the power under Rule 26(b) to “limit” discovery.<sup>83</sup> In the event of abuse, the court can exercise “more stringent control” over discovery<sup>84</sup> and can order sanctions.<sup>85</sup> Sanctions are appropriate when a party acts “in bad faith, vexatiously, wantonly, or for oppressive reasons.”<sup>86</sup>

Certainly, judges have wide discretion to act in managing cases,<sup>87</sup> and amendments to the federal discovery rules since 1983 have led toward greater exercise of judicial management and judicial discretion.<sup>88</sup> But active case management does not mean unfettered discretion. For example, the Advisory Committee note to Rule 16 cautions: “The rule does not provide the court with authority to enter such a case-management or other order [related to privilege protections] without party agreement . . . .”<sup>89</sup> Much scholarly debate focuses on the benefits and challenges associated with judicial discretion and the optimal level of such discretion in managing the pre-trial process.<sup>90</sup> Allowing judges some discretion in managing the pre-trial process is valuable, of course, as it obviates the need for more detailed Rules and allows for flexibility to tailor discovery to the needs and situation of each particular case.<sup>91</sup>

Excessive reliance on judicial discretion in managing discovery can be problematic for several reasons. First, it can lead to disparate treatment of similarly situated parties,<sup>92</sup> despite the fact that the system seeks uniform treatment from one courtroom to the next. Second, it provides opportunity for judges to pursue personal agendas.<sup>93</sup> Finally, discovery rulings by judges are rarely appealed. When rulings are appealed, parties must show “abuse of discretion,” which standard is rarely shown.<sup>94</sup> Understanding that judges may have widely varying approaches to discovery, parties may engage in forum shopping, “as litigants seek districts that contain judges whose views on discovery suit their particular needs.”<sup>95</sup>

## V. Why Courts Should Not Mandate Use of Predictive Coding

The foregoing Parts of this paper described predictive coding technology, introduced current cases that bring the issue to the fore, and discussed the respective roles of the parties and the court. This Part seeks to assimilate that information to address the ultimate question: should courts mandate the use of predictive coding in electronic discovery over a party’s objection? Analysis suggests that courts should not intrude in parties’ decisions on whether to employ predictive coding in electronic discovery, for two broad reasons. The first reason relates to legitimate concerns that parties and attorneys may have about predictive coding, most notably a concern about privilege protection. The second reason relates to respect for the traditional judicial role.

## A. Parties and Attorneys May Have Legitimate, Good Faith Reasons to Prefer Keyword Searching

Certain proffered reasons that parties and attorneys may be reluctant to adopt predictive coding—lack of transparency of the process, concern about accuracy of the results, and effect on privilege protection<sup>96</sup>—cannot lightly be dismissed. Keyword searching is an established, judicially recognized method.<sup>97</sup> Empirical data establishing the superiority of predictive coding is sparse.<sup>98</sup> Parties may reasonably have doubts about predictive coding. These facts suggest that parties can in good faith choose keyword searching as a method of culling relevant documents, to fulfill their obligation to produce relevant, responsive documents, and to meet the overall goal of “just, speedy, and inexpensive” discovery set by Rule 1.<sup>99</sup>

Note that attorneys have a legal and ethical obligation to make sure that discovery is conducted properly. Rule 26(g) requires that an attorney responding to a discovery request must certify that he or she has made a “reasonable inquiry”<sup>100</sup> and that the response is “complete and correct.”<sup>101</sup> The attorney’s responsibility cannot be abrogated by judicial rulings. These obligations imply that attorneys have a level of comfort with and choice regarding methods used. They further presume party control of the process. For judges to mandate use of particular technology in which the attorney does not have confidence would put attorneys in an untenable bind, inconsistent with the spirit of the system.

Parties have an obligation to avoid abusing the discovery process through such tactics as gamesmanship, courtesy, or failure to participate in discovery in an orderly fashion.<sup>102</sup> But there is little indication that parties seek to employ keyword searching instead of predictive coding as a litigation tactic—to harass, delay, or impose costs on the other side. It appears, rather, that parties reasonably and in good faith disagree about the most appropriate technology to be employed.<sup>103</sup> In fact, if proponents’ assurances about the cost savings of predictive coding are accurate, parties choosing to employ keyword searching instead are imposing higher costs on themselves.

One might debate the presence of abuse or failure to achieve “just” discovery if it could be shown that keyword searching were chosen by a producing party specifically out of a belief that the method would more likely return over- or under-inclusive results. Either extreme is a problem. Excess production imposes a burden on the requesting party, as it increases the burdens of examining the documents; under-production means that relevant and important documents may not be produced. But the empirical data, at this point, do not conclusively show that predictive coding is more accurate than keyword searching.<sup>104</sup> And if particular keyword searches are inadequate, the court can order that party to refine or expand the search as appropriate.<sup>105</sup> Those remedies are more appropriate than taking the extraordinary step of mandating use of predictive coding.

Parties may also reasonably resist use of predictive coding methods due to concerns about protection of the attorney-client privilege. In cases involving discovery of ESI, regardless of the search method used, when enormous numbers of documents are produced, the inadvertent production of privileged documents is likely.<sup>106</sup> The Rules so recognize and provide for mechanisms to handle this problem.<sup>107</sup> Ideally, in their pre-trial discovery plan, the parties will address what should happen in such a case.<sup>108</sup> Parties may make the arrangement of their choosing, but typically will agree to a “clawback” arrangement.<sup>109</sup> This sort of agreement is codified in Rule 26(b)(5), which provides that if privileged information is produced in discovery, then the producing party so notifies the receiving party, who must “promptly return, sequester, or destroy the specified information and any copies that it has; [and] must not use or disclose the information until the claim is resolved.”<sup>110</sup> Privileged documents that are clawed back cannot be used as evidence in the case.<sup>111</sup>

Recall that one of the major arguments for use of predictive coding is the promise of cost savings, and that cost savings through predictive coding are expected primarily because pre-production privilege review is eliminated or greatly curtailed.<sup>112</sup> If a party thus hopes to achieve the cost benefits of this method, that party accepts a greater risk of inadvertent production of privileged documents. Privilege protection would not be lost, due to the availability of clawbacks and the provisions of Federal Rule of Civil Procedure 26(b) and Federal Rule of Evidence 502.<sup>113</sup>

The problem with clawback arrangements, of course, is that “the bell cannot be unrung.”<sup>114</sup> Even though the opponent cannot offer the privileged documents as evidence in the case, “it will have seen it, likely be unable to forget it, and be able to use the information to strategize for its case.”<sup>115</sup> Furthermore, clawbacks do not prevent use of the privileged materials by third parties.<sup>116</sup> Parties considering the use of predictive coding would do well to weigh these risks.

Given these significant risks, courts should not *compel* parties to employ predictive coding, which method normally foregoes a final, pre-production privilege review and therefore, puts the party at a greater risk of disclosing privileged documents. The attorney-client privilege is “the oldest of privileges for confidential communications known to the common law” and serves the critical purpose of “encourag[ing] full and frank communication between attorneys and their clients and thereby promote[s] broader public interests in the observance of law and administration of justice.”<sup>117</sup> Federal Rule of Evidence 502 demonstrates Congress’ strong intent to protect the privilege.<sup>118</sup> While under Rule 502 any privileged documents could be clawed back,<sup>119</sup> clawback provides insufficient protection. The important policy behind protection of the attorney-client privilege outweighs the hope that use of predictive coding would provide cost savings and perhaps more accurate production of ESI. The privilege is owned by the client, “who alone may waive it.”<sup>120</sup> The client may rightly object to any method that contemplates foregoing a final, pre-production privilege review.<sup>121</sup>

## B. Judges Do Not Have Sufficient Reason to Intervene in Parties' Decisions on Search Technology

While Rule 26(b) explicitly grants courts the power to "limit" discovery,<sup>122</sup> case law shows that courts have imposed "limitations" to address proportionality problems, in cases where the requested discovery was overly burdensome or costly, given the nature of the case or the value of the information likely to be gained.<sup>123</sup> This is consistent with the Advisory Committee notes accompanying this Rule.<sup>124</sup> Orders to limit the number of interrogatories, limit the universe of ESI to be examined, or phase discovery<sup>125</sup> fit the plain meaning of the word "limit."<sup>126</sup> Ordering a party to employ a *different* technology to search its own ESI would be unlike those sorts of limitations and does not go to a concern about proportionality. Such an order would not naturally be regarded as a "reduction" or "restriction" of discovery.

Certainly, courts have discretion to act, apart from the specific grants of power in the Rules.<sup>127</sup> But courts generally have intervened in parties' discovery activities only in the case of abuse—gamesmanship, courtesy, or failure to participate in discovery as required.<sup>128</sup> As noted above, parties' motivation to prefer keyword searching appears not to be for nefarious reasons. Absent a showing of abuse, therefore, courts should hesitate to intervene. As one scholar bluntly and correctly notes, "No matter how talented the judge, determinations about the proper use of discovery and specific tools are better left to the attorneys on the case, who are more familiar with the facts and needs of the case, and who are directly responsible to their clients."<sup>129</sup>

Certain judges may have a keen interest in, and deep understanding of, predictive coding. Perhaps understandably, they may wish to use their rulings to further acceptance of a nascent, promising technology. This, however, gives rise to a serious concern about disparate treatment from one court to the next as well as the possibility of forum shopping.<sup>130</sup> Moreover, even if predictive coding has the potential to be more effective than keyword searching and will be widely used and acclaimed in the future, is it the place of a judge to foster use of particular technology through court rulings? Parties still would have reason for concern about privilege protection.<sup>131</sup> Mandating particular technology still would be a departure from the long-standing tradition of party-controlled discovery.<sup>132</sup> Those concerns should override any incremental benefits of predictive coding over keyword searching.<sup>133</sup>

A broader question may be whether a judge or indeed any government actor should or can effectively promote public acceptance of new technology. Government efforts to promote such technologies as electric cars and compact fluorescent light bulbs are well known.<sup>134</sup> The analogy is imperfect: a judge's effort to mandate use of predictive coding in discovery because the technology is thought to be more efficient and effective differs in important ways from legislative or executive attempts to support technology because it is thought to be more environmentally friendly.<sup>135</sup> But the literature assessing the latter type of government efforts may offer some relevant lessons.

Most obviously, that literature emphasizes the difficulty of such endeavors and the complexity of issues involved.<sup>136</sup> The literature also notes potential obstacles to technological change, among them "industry opposition, consumer reluctance to purchase unfamiliar new technologies, . . . [and] the rapidly changing technology landscape that may cause frequent rethinking of the optimal technology pathways."<sup>137</sup> The literature notes government's limited success in identifying the best technology.<sup>138</sup> Finally and most importantly, there is the following conclusion: "The consumer is the ultimate arbiter of most technological changes. If consumers are unwilling to accept or pay for a new technology, that technology is unlikely to prosper. Therefore, policies that attempt to 'push' a technology onto unreceptive or even uninterested consumers are particularly prone to fail."<sup>139</sup> Surely a corollary of this last point is that truly effective and useful new technologies are likely to be accepted by consumers, without the need for government mandates.

One scholar recommends that rather than mandating new technologies, government "should identify what are the key goals or problems it is trying to address and then not discriminate against any technologies that can help achieve those stated objectives."<sup>140</sup> This approach, in essence, would align with the traditional approach of the Federal Rules—allowing parties to direct discovery, including selecting search technologies.

## VI. Conclusion

The aim of this paper is not to denigrate technological progress. Nor is it to denigrate predictive coding—a promising method for culling relevant documents in electronic discovery. If parties have confidence in and choose to use that method, courts should accept it as a reasonable method.

Rather, the aim of this paper is to provide an objective look at predictive coding and to offer a reminder of the time-tested underpinnings of the discovery process under the Federal Rules of Procedure. The system works best as a party-controlled system, with courts overseeing the process to control abuse by the parties, ruling on such things as time, place, and subject matter for discovery. While judges in recent years have taken a more aggressive approach to case management, stepping too far outside the traditional judicial role can lead to the appearance of unfairness and even charges of bias that undermine the system.

This paper examined whether judges should intervene in parties' decisions on what type of search technology to use in electronic discovery. Recent cases suggest the possibility that judges could soon mandate use of predictive coding technology over a party's objections. This issue implicates important policy considerations that go to the heart of the American adversarial system. Such a mandate would be unwise. Parties and attorneys may be resistant to that technology

for legitimate reasons, including lack of transparency of the process, questions about whether this method is more accurate than the more familiar keyword searching, and heightened risk of inadvertent disclosure of privileged materials, which can provide strategic advantages to the opponent. Moreover, such a mandate would depart from the traditional judicial role without adequate justification.

Mandating use of predictive coding in electronic discovery, therefore, would be an ill-advised judicial intrusion. Judges should not issue such mandates. This is not an issue that requires a change in the Rules; it simply requires the exercise of judicial restraint. Groups such as the Sedona Conference and the Seventh Circuit Electronic Discovery Program could provide welcome leadership in advocating this sort of restraint. These groups might also encourage greater transparency regarding predictive coding products and services, development of empirical data regarding its effectiveness, and further improvement in predictive coding technology, which may lead to wider voluntary use of that technology and perhaps more “just, speedy, and inexpensive discovery.”

## FOOTNOTES

---

<sup>1</sup> HENRY DAVID THOREAU, WALDEN 25 (W.W. Norton & Co. 1966) (1854).

<sup>2</sup> The term “electronically stored information” is not defined in the Federal Rules of Civil Procedure, although it is used there. *See infra*, notes 64-76. In Notes accompanying the 2006 Amendments to those Rules, the Judicial Conference Committee stated that the term “is expansive and includes any type of information that is stored electronically” and “is intended to be broad enough to cover all current types of computer-based information, and flexible enough to encompass future changes and developments.” Fed. R. Civ. P. 34(a)(1) advisory committee’s note. In most companies, the vast majority of documents are created and stored electronically. *On Its Way, At Last: No Longer a Joke, the “Paperless Office” Is Getting Closer*, ECONOMIST, Oct. 9, 2008, at 86.

<sup>3</sup> See, e.g., Gretchen Morgenson, *Case on Mortgage Official Is Said to Be Dropped*, N.Y. TIMES, Feb. 20, 2011, at A20 (reporting closure of criminal case against Countrywide Financial’s chief executive but noting that internal emails calling the company’s loans “toxic” and “poison” were likely to be pivotal in civil cases against the company); Landon Thomas, Jr., *Prosecutors Build Bear Stearns Case on E-Mails*, N.Y. TIMES (June 20, 2008), <http://www.nytimes.com/2008/06/20/business/20bear.html?pagewanted=all> (describing prosecutors’ use of hedge fund executive’s internal email stating “I think we should close the funds now,” while the firm “presented a sunny picture to worried investors”); Steve Lohr, *In an Antitrust Suit, a Tiny Ex-Partner Is Taking Aim at Microsoft*, N.Y. TIMES (May 31, 1999), <http://www.nytimes.com/1999/05/31/business/in-an-antitrust-suit-a-tiny-ex-partner-is-taking-aim-at-microsoft.html?pagewanted=all&src=pm> (noting that antitrust case against Microsoft was largely built on internal emails sent by Bill Gates and other executives); Ravi Mandalia, *Google Denied Exclusion of Lindholm Email from Oracle Lawsuit*, ITPROPORTAL (Feb. 8, 2012), <http://www.itproportal.com/2012/02/08/google-denied-exclusion-of-lindholm-email-from-oracle-lawsuit> (reporting that Federal Circuit’s ruling to allow use of incriminating email was “a potentially damaging blow” in billion dollar patent lawsuit); “*Smoking Gun*” *Emails Released in 2009 Buffalo Plane Crash*, ABC NEWS (Oct. 21, 2011), <http://abcnews.go.com/blogs/headlines/2011/10/smoking-gun-emails-released-in-lawsuit-over-2009-buffalo-plane-crash>.

<sup>4</sup> See JOHN F. GANTZ ET AL., INT’L DATA CORP., THE DIVERSE AND EXPLODING DIGITAL UNIVERSE: AN UPDATED FORECAST OF WORLDWIDE INFORMATION GROWTH THROUGH 2011 3 (2008) (projecting a ten-fold increase in ESI from 2006-11). Documents responsive to discovery requests may be found in such locations as servers, desktop and laptop computers at work and at employees’ homes, on cloud-based systems, on Voice over Internet protocol (VOIP) systems, on removable media such as flash drives, and on devices such as smart phones. This proliferation of technologies accounts in part for the proliferation of data and complicates the task of retrieving and evaluating ESI. See, e.g., Mia Mazza et al., *In Pursuit of FRCP 1: Creative Approaches to Cutting and Shifting the Costs of Discovery of Electronically Stored Information*, 13 J. RICH. J.L. & TECH. 11, ¶ 1 (2007); Douglas Oard et al., *Evaluation of Information Retrieval for E-Discovery*, 18 ARTIFICIAL INTELL. L. 347, 350-51 (2010). Various laws compelling retention of data also contribute to the proliferation. See, e.g., 17 C.F.R. § 210.2-06 (2003) (implementing Sarbanes-Oxley Act and requiring retention of audit-related records, including electronic records, for seven years); 45 C.F.R. § 164.530(j)(2) (implementing Health Insurance Portability and Accountability Act of 1996 and requiring retention of medical records for six years).

<sup>5</sup> The Judicial Conference Advisory Committee, tasked with revising the Federal Rules of Civil Procedure to address electronic discovery, lists particular challenges:

Electronically stored information is characterized by exponentially greater volume than hard-copy documents. Commonly cited current examples of such volume include the capacity of large organizations’ computer networks to store information in terabytes, each of which represents the equivalent of 500 million typewritten pages of plain text, and to receive 250 to 300 million e-mail messages monthly. Computer information, unlike paper, is also dynamic; merely turning a computer on or off can change the information it stores. Computers operate by overwriting and deleting information, often without the operator’s specific direction or knowledge. A third important difference is that electronically stored information, unlike words on paper, may be incomprehensible when separated from the system that created it.

---

JUDICIAL CONFERENCE ADVISORY COMM., THE NEW E-DISCOVERY RULES 9 (2d ed. 2009), available at [http://www.legalpub.com/\\_images/sample-pages/2009-ediscovery.pdf](http://www.legalpub.com/_images/sample-pages/2009-ediscovery.pdf).

<sup>6</sup> The costs of electronic discovery are generally borne by the producing party, but may be shifted at least in part to the requesting party when “good cause” is shown. See Hon. James C. Francis IV, *Cost Shifting in E-Discovery*, in MANAGING E-DISCOVERY AND ESI 591, 603-05 (Michael D. Berman et al. eds., 2011); Robert Hardaway et al., *E-Discovery’s Threat to Civil Litigation: Reevaluating Rule 26 for the Digital Age*, 63 RUTGERS L. REV. 521, 566-73 (2011). In addition to concerns about costs of discovery, there are issues of legal liability for improper handling of ESI, such as its improper destruction when litigation is anticipated, see Burke T. Ward et al., *Recognizing the Impact of E-Discovery Amendments on Electronic Records Management*, 26 INFO. MGMT. 350, 352 (2009), or improper disclosure of confidential information. See, e.g., Robert Peglar, *Evidence Management Solutions for Mitigating E-Records Risk*, 41 INFO. MGMT. J. 56, 57-58 (2007). Many sectors of society are dealing with challenges in connection with the growing amount of ESI. See *Data, Data Everywhere*, ECONOMIST, Feb. 25, 2010, at 53 (citing challenges for science, government, and the arts). Corporate emails are relevant evidence in the accounting audit process, and therefore, auditors face challenges in culling the critical emails, sometimes intended to deceive those who may later read these emails, from a large mass of information. See Roger S. Debreceny & Glen L. Gray, *Data Mining of Electronic Mail and Auditing: A Research Agenda*, 25 J. INFO. SYSTEMS 195 (2011) (describing data mining tools used in auditing, including keyword searching).

<sup>7</sup> See Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 254-55, 260 n. 9 (D. Md. 2008) (describing discovery protocols); Oard et al., *supra* note 4, at 350-53 (2010) (describing process of discovery employing keyword searches); Sedona Conf., *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 196, 200-02 (2007).

<sup>8</sup> Da Silva Moore v. Publicis Groupe, 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Feb. 24, 2012).

<sup>9</sup> Kleen Prods., LLC v. Packaging Corp. of Am., Civil Case No. 1:10-cv-05711 (N.D. Ill. filed Sept. 9, 2010).

<sup>10</sup> See Order Approving the Use of Predictive Coding for Discovery, Global Aerospace, Inc. v. Landow Aviation, L.P., Consolidated Case No. CL61040 (Loudon Cnty. April 23, 2012). The court thus approved the defendant’s use of predictive coding over the plaintiff’s objection. Basic facts about the dispute can be found in the plaintiff’s brief. See Brief in Opposition to Motion for Protective Order Regarding Electronic Documents and “Predictive Coding” at 2, Global Aerospace, Inc. v. Landow Aviation, L.P., Consolidated Case No. CL61040 (Loudon Cnty. April 16, 2012) (arguing that “There is certainly no reason that Virginia law permits such a radical departure from the standard practice of human review of documents.”). In this case, the Circuit Court for Loudoun County, Virginia, applied the Virginia Rules of Civil Procedure. This paper will focus on the Federal Rules of Civil Procedure. Treatment of predictive coding under each state’s rules is beyond the scope of this paper, but to the degree that a particular state’s discovery rules are modeled on the Federal Rules, the same analysis and concerns would apply. A recent study showed that 24 of the 29 states that have ESI-specific rules emulate the federal rules. See Robert Dale Klein & Hon. Joseph F. Murphy, Jr., *The States: What Are the Laboratories of Federalism Doing?*, in MANAGING E-DISCOVERY AND ESI 49, 50 (Michael D. Berman et al. eds., 2011).

<sup>11</sup> See, e.g., Matthew Nelson, *Federal Judges Consider Important Issues that Could Shape the Future of Predictive Coding Technology*, FORBES (Feb. 20, 2012, 1:42 p.m.), <http://www.forbes.com/sites/benkerschberg/2012/02/20/federal-judges-consider-important-issues-that-could-shape-the-future-of-predictive-coding-technology/>; Matthew Nelson, *Plaintiffs Ask Judge Nan R. Nolan to Go Out on a Limb in Kleen Products Predictive Coding Case*, E-DISCOVERY 2.0 (April 13, 2012, 8:58 a.m.), <http://electronicdiscovery.info/plaintiffs-ask-judge-nan-r-nolan-to-go-out-on-a-limb-in-kleen-products-predictive-coding-case-electronic-discovery/> (calling plaintiffs’ request that the judge order defendant to use predictive coding “shocking to many observers”); Michael Roach, *Predictive Coding Watch: “Kleen Products” in Illinois*, EDD UPDATE (April 17, 2012, 1:25 p.m.), <http://www.eddupdate.com/2012/04/predictive-coding-watch-kleen-products-in-illinois.html>; Judy Selby, *New Horizons in Three Concurrent Cases: Predicting the Future of Predictive Coding*, DISCOVERY ADVOCATE (April 25, 2012), <http://www.discoveryadvocate.com/2012/04/25/new-horizons-in-three-concurrent-cases-predicting-the-future-of-predictive-coding/>; Dave Walton, *Technology: Da Silva Moore Sets Stage for Predictive Coding’s Advancement*, INSIDE COUNSEL (April 30, 2012), <http://www.insidecounsel.com/2012/04/30/da-silva-moore-sets-stage-for-predictive-codings-a>.

<sup>12</sup> See, e.g., *Victor Stanley*, 250 F.R.D. at 254-55 (noting that parties employed five pages of keyword search terms). Douglas Oard, Professor in the College of Information Studies at the University of Maryland, and his co-authors more precisely define the methods that may be employed in keyword searching:

The term “keyword searching” . . . has been used in the IR [Information Retrieval] literature to refer to any or all of exact string matching, substring matching, Boolean search, or statistical ranked retrieval, applied to any or all of free text terms (e.g., space-delimited tokens or character n-grams), manually or automatically assigned controlled vocabulary terms, with or without augmentation by any combination of stemming, wildcards, multi-word phrase formation, proximity and/or word order restrictions, field restrictions, and/or a variety of other operators.

Oard et al., *supra* note 4, at 359-60.

---

<sup>13</sup> See, e.g., Osborne v. C.H. Robinson Co., 2011 U.S. Dist. LEXIS 123168 (N.D. Ill. Oct. 25, 2011). For example, they may decide to search only files of certain people or to search only files developed during certain time frames.

<sup>14</sup> See, e.g., Alexandra Hanson, *Legal Outsourcing to India: So Hot Right Now!*, 62 SMUL. REV. 1889 (2009) (discussing outsourcing of document review and legal research); Aaron R. Harmon, *The Ethics of Legal Process Outsourcing—Is the Practice of Law a “Noble Profession or Is It Just Another Business?*, 13 J. TECH. L. & POL’Y 41 (2008); Brandon Robers, Current Development, *The Firm Is Flat: Ethical Implications of Legal Offshoring*, 23 GEO. J. LEGAL ETHICS 700 (2010). Outsourcing may provide cost savings and efficiency, but may present challenges of confidentiality and quality control. It may also be controversial at a time when U.S. law jobs are being eliminated. E.g., Hanson, *supra*, at 1897-907.

<sup>15</sup> The issue of clawback of inadvertently produced privileged documents, depicted as the last step in Figure 1, is discussed *infra*, text accompanying notes 107-16.

<sup>16</sup> See, e.g., Custom Hardware Eng’g & Consulting, Inc. v. Dowell, 2012 U.S. Dist. LEXIS 146, at \*6 (E.D. Mo. Jan. 3, 2012); Jason R. Baron & Michael D. Berman, *Designing a “Reasonable” E-Discovery Search: A Guide for the Perplexed*, in MANAGING E-DISCOVERY AND ESI 479, 482 (Michael D. Berman et al. eds., 2011) (“Automated methods of search, starting with applying more sophisticated approaches to keyword searching, are the wave of the future.”); Sedona Conf., *supra* note 7, at 200. Keyword searching has been shown to yield more complete and accurate production than old-fashioned, manual review of documents. See, e.g., Patrick Oot et al., *Practitioners’ View: Mandating Reasonableness in a Reasonable Inquiry*, 87 DENV. U. L. REV. 533, 548-52 (2010). In any event, manual searching of vast quantities of data would be impracticable.

<sup>17</sup> See, e.g., *Osborne*, 2011 U.S. Dist. LEXIS 123168, at \*12.

<sup>18</sup> See, e.g., *Custom Hardware*, 2012 U.S. Dist. LEXIS 146, at \*8-12; Helmert v. Butterball, LLC, 2010 U.S. Dist. LEXIS 60777 (E.D. Ark. May 27, 2010); William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 135 (S.D.N.Y. 2009). Whether judges are well-equipped to craft or evaluate particular search terms is questioned. See, e.g., Eurand, Inc. v. Myland Pharm, Inc., 266 F.R.D. 79, 84 (D. Del. 2010) (“Neither lawyers nor judges are generally qualified to opine that certain search terms or files are more or less likely to produce information than those keywords or data actually used or reviewed.”); Marian Riedy et al., *Mediated Investigative E-Discovery*, 2010 FED. CTS. L. REV. 79, 84 (2010) (“Analyzing keyword searches is a strain on judicial resources, and may also be outside the ken of judicial abilities. Courts now evaluate keyword search proposals just as they would document requests . . . .”).

<sup>19</sup> See, e.g., *Custom Hardware*, 2012 U.S. Dist. LEXIS 146, at \*6; *Osborne*, 2011 U.S. Dist. LEXIS 123168, at \*4; William A. Gross, 256 F.R.D. at 135; Baron & Berman, *supra* note 16, at 482. Crafting effective search terms is challenging in part because it may be difficult to capture all synonyms that express a relevant idea, and also because ESI may include many informal documents such as emails or text messages, where misspellings, slang, or acronyms are often used. David D. Cross & Sanya Sarich Kerksiek, *Using Electronic Search Tools and Search Methodology Experts in E-Discovery: A Discussion of Recent Case Law and Other Authorities*, in MANAGING E-DISCOVERY AND ESI 439, 443-46 (Michael D. Berman et al. eds., 2011); Lori Heilman, Comment, *Federal Courts’ Reactions to Inadequate Keyword Searches: Moving Toward a Predictable and Consistent Standard for Attorneys Employing Keyword Searches*, 78 U. CIN. L. REV. 1103, 1107 (2010). In contrast, crafting effective keyword searches for LEXIS or Westlaw may be an easier task, as legal opinions and law review articles are more formal in nature and may employ more standard and predictable terminology.

<sup>20</sup> Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 262 (D. Md. 2008); Lauren Katz, Current Development, *A Balancing Act: Ethical Dilemmas in Retaining E-Discovery Consultants*, 22 GEO. J. LEGAL ETHICS 929 (2009) (noting that to fulfill ethical duties of diligence and competence, attorneys may need to obtain assistance of e-discovery expert).

<sup>21</sup> See, e.g., *Victor Stanley*, 250 F.R.D. at 262; Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008); Cross & Kirksiek, *supra* note 19, at 456-60; Oard et al., *supra* note 4, at 351 (noting “emergence of a body of case law questioning the ad hoc, unexplained, and/or unilaterally deployed use of single keywords as search terms”); Riedy et al., *supra* note 18, at 84. See also Sedona Conf., *supra* note 7, at 212 (“Parties should expect that their choice of search methodologies will need to be explained . . . .”).

<sup>22</sup> Predictive coding is a type of “concept searching.” Professor Oard and his co-authors describe the variety of methods that may be employed in concept searching:

Controlled vocabulary indexing (manual or automatic, with or without thesauri), multi-word phrase formation (by statistical and/or linguistic means), statistical query expansion methods, knowledge representation languages and inference systems from artificial intelligence, unsupervised learning approaches (including term clustering, document clustering, and factor analytic methods such as latent semantic indexing), as well as simple stemming, wildcards, spelling correction and string similarity measures.

Oard et al., *supra* note 4, at 360. See also Venkat Rangan, Discovery of Related Terms in a Corpus Using Reflective Random Indexing (July 28, 2011) (unpublished paper, presented at SIGIR 2011 Information Retrieval for E-Discovery Workshop, Beijing, China), available at <http://www.umiacs.umd.edu/~oard/desi4/papers/rangan.pdf> (describing various concept searching methods).

---

<sup>23</sup> NICHOLAS M. PACE & LAURA ZAKARAS, RAND INST. FOR CIVIL JUSTICE, WHERE THE MONEY GOES: UNDERSTANDING LITIGANT EXPENDITURES FOR PRODUCING ELECTRONIC DISCOVERY 59 (2012) [hereinafter RAND REPORT]. See also Philip Cohen & Lauren Harrison, *Predictive Coding Is a New Tool in the E-Discovery Toolbox*, N.Y.L.J. (Mar. 19, 2012), <http://www.newyorklawjournal.com/PubArticleFriendlyNY.jsp?id=1202545983136&slreturn=1> (describing predictive coding process).

<sup>24</sup> RAND REPORT, *supra* note 23, at 60. “[T]he best results will be achieved if the attorneys most closely involved in the case select the seed documents and review sampled extracts . . . . Moreover, attorney judgment continues to loom large in the process after the application has completed its work, with eyes-on review required, for example, to check documents of unknown relevance and responsiveness or look for privileged communications.” *Id.* at 61.

<sup>25</sup> The issue of clawback of inadvertently produced privileged documents, depicted as the last step in Figure 2, is discussed *infra*, text accompanying notes 107-16.

<sup>26</sup> RAND REPORT, *supra* note 23, at 60. Because attorney “review consumed at least 70 percent of the total costs of document production [in at least half of the studied cases], this single area, described by one participant in our study as the ‘black hole’ of the entire process, is an obvious target for reducing e-discovery expenditures.” RAND REPORT, *supra* note 23, at 41.

The patent application for one predictive coding method contemplates little or no human review of the documents produced by the software: “[T]he Invention can limit, or avoid altogether, the final human review of the computer files before production. This allows for very significant cost savings by responding parties by reduction of review times with protections from waiver of rights by inadvertent disclosure of confidential privileged [sic] provided by ‘Confidentiality,’ ‘Clawback,’ and/or ‘Quickpeek’ agreements, and the terms and protection offered by newly enacted Rule 26(b)(5)(B), Federal Rules of Civil Procedure, and newly enacted Rule 502, Federal Rules of Evidence, and orders entered there-under.” Patent Application 20100205020 at [36]. The issue of privilege protection is discussed *infra*, text accompanying notes 107-16.

<sup>27</sup> RAND REPORT, *supra* note 23, at 67 (noting difficulty of obtaining accurate information).

<sup>28</sup> *Id.*

<sup>29</sup> See Systems and Methods for Predictive Coding, U.S. Patent No. 7933859 (filed May 25, 2010) (issued April 26, 2012) (assigned to Recomind, Inc.); Identifying Related Objects Using Quantum Clustering, U.S. Patent No. 8010534 (filed Aug. 31, 2007) (issued Aug. 30, 2011) (assigned to OrcaTec, LLC); Method and Apparatus for Processing Electronically Stored Information for Electronic Discovery, U.S. Patent 8090705 (filed Sept. 15, 2008) (issued Jan. 3, 2012) (assigned to Symantec Corp.).

<sup>30</sup> See, e.g., Predictive Coding of Documents in an Electronic Discovery System, U.S. Patent Application 20100250474 (filed March 24, 2010) (published Sept. 30, 2010); System and Method for Establishing, Managing, and Controlling the Time, Cost, and Quality of Information Retrieval and Production in Electronic Discovery, U.S. Patent Application 20100205020 (filed Feb. 9, 2009) (published Aug. 12, 2010); Identifying Related Objects Using Quantum Clustering, U.S. Patent Application 20110295856 (filed Aug. 8, 2011) (published Dec. 1, 2011); Systems and Methods for Electronic Document Review, U.S. Patent Application 20100077301 (filed Sept. 21, 2009) (published March 25, 2010).

<sup>31</sup> See, e.g., E-mail from Herbert L Roitblat, Chief Scientist and Chief Technology Officer, OrcaTec, LLC (May 14, 2012, 10:12 a.m. EDT) (on file with author). Many companies purport to offer predictive coding services. In addition to those mentioned *supra* note 29, other providers include Clearwell Systems, Inc., Equivio, Ltd., Applied Discovery, Inc., and Oracle Corp., as well as law and accounting firms. See *supra*, note 22, for listing of various concept searching technologies that might be employed. The author takes no position on the effectiveness of any particular provider’s services or products, and certainly does not intend to denigrate any particular provider.

<sup>32</sup> See, e.g., Joe Dysart, *A New View of Review: Predictive Coding Vows to Cut E-Discovery Drudgery*, ABA J. (Oct. 1, 2011), [http://www.abajournal.com/magazine/article/a\\_new\\_view\\_of\\_review\\_predictive\\_coding\\_vows\\_to\\_cut\\_e-discovery\\_drudgery/](http://www.abajournal.com/magazine/article/a_new_view_of_review_predictive_coding_vows_to_cut_e-discovery_drudgery/); John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES, March 5, 2011, at A1 (noting that concept searching “adds an inferential layer of analysis, mimicking the deductive powers of a human Sherlock Holmes.”); Matthew Nelson, *2012: Year of the Dragon – and Predictive Coding. Will the eDiscovery Landscape Be Forever Changed?*, E-DISCOVERY BLOG (Jan. 23, 2012, 5:00 a.m.), <http://www.clearwellsystems.com/e-discovery-blog/2012/01/23/year-of-the-dragon-and-ediscovery-predictive-coding/>. The Conclusion of the RAND REPORT is more measured, calling predictive coding “[t]he most promising alternative available today,” but noting numerous ongoing challenges. RAND REPORT, *supra* note 23, at 97.

<sup>33</sup> RAND REPORT, *supra* note 23, at xviii (“Because this is nascent technology, there is little research on how the accuracy of predictive coding compares with that of human review.”). Developing such studies is a daunting task, due to costs, difficulty in accessing documents and information, and other challenges. *Id.* at 3-4. One notable study was undertaken by TREC (Text Retrieval Conference) Legal Track in 2006-09, using tobacco litigation electronic documents released under the Tobacco Master Settlement Agreement as the test collection to compare “statistical ranked retrieval techniques” with methods based on “Boolean, proximity, and truncation operators.” See Oard et al., *supra* note 4, at 366-79. While the study concludes that the former techniques are promising, the authors note the need for much more empirical research. *Id.* at 377. See also Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More*

---

*Efficient than Exhaustive Manual Review*, 17 RICH. J.L. & TECH. 11 (2011); Herbert L. Roitblat et al., *Document Classification in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 6 J. AM. SOC. INFO. SCI. & TECH. 70 (2010) (finding superior results from unspecified “computer-aided systems” as compared with manual review of documents). Note that state of the art is no longer manual review. *See supra*, text accompanying note 16. Note further that these studies do not specify precisely the computer-assisted methods that were employed, which limits their usefulness.

<sup>34</sup> See Oard et al., *supra* note 4, at 360. Interestingly, these authors observe that despite the widespread use of keyword searching techniques, the term “keyword searching” is “used almost exclusively in a derogatory fashion, to refer to any method which an author believes is inferior to their preferred technique,” while ““concept searching’ is almost uniformly used with a positive connotation, in both technical and marketing literature.” *Id.*

<sup>35</sup> See Oard et al., *supra* note 4, at 377. Judges have also noted this phenomenon. *See, e.g.*, Equity Analytics LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008) (“[L]awyers state as facts what are actually highly debatable propositions as to the efficacy of various methods used to search electronically stored information.”).

<sup>36</sup> *Id.* at 378.

<sup>37</sup> RAND REPORT, *supra* note 23, at xviii (“[A]t the end of 2011, we could find no evidence in the published record that any vendor, law firm, or litigant had used predictive coding in a publicized case that named the parties and court jurisdiction.”).

<sup>38</sup> *See, e.g.*, R. Eric Hutz, *E-Discovery: Using Predictive Coding to Manage E-Discovery Costs and Risks*, INSIDE COUNSEL (Feb. 23, 2012), <http://www.insidecounsel.com/2012/02/23/e-discovery-using-predictive-coding-to-manage-e-di> (noting that many lawyers view predictive coding as a “black box” process” using “complex computer algorithms . . . not easily understood or explainable by anyone other than a computer scientist”). More ominously, some have wondered whether resistance to predictive coding may stem in part from lawyers who would stand to lose revenue long derived from manual review of documents. RAND REPORT, *supra* note 23, at 76.

<sup>39</sup> *See* Nelson, *supra* note 32 (“[I]t is difficult to defend a technological process that isn’t always clear . . . In short, using black box technology that is difficult to use and understand is perceived as risky, and many attorneys have taken a wait-and-see approach because they are unwilling to be the guinea pig.”).

<sup>40</sup> *Id.*; RAND REPORT, *supra* note 23, at 74-76 (noting a particular concern about under-inclusion—a failure to identify and produce responsive documents, especially highly technical or non-text documents); Hutz, *supra* note 38. *See also generally* Hardaway et al., *supra* note 6, at 551-64 (noting problems with both keyword searching and concept searching).

<sup>41</sup> *See, e.g., id.* at 77-79; Cohen & Harrison, *supra* note 23; Hutz, *supra* note 38 (“There is an open question whether predictive coding is sufficiently rigorous and transparent that a court would be satisfied that its use [meets reasonableness requirements of Federal Rules of Civil Procedure].”); Nelson, *supra* note 32 (arguing that recent “improvements in transparency will simplify the user experience and improve accuracy which will reduce longstanding concerns about defensibility”).

<sup>42</sup> *See* Hutz, *supra* note 38. This concern is discussed at some length, *infra* notes 107-16 and accompanying text.

<sup>43</sup> Sedona Conf., *supra* note 7, at 204.

<sup>44</sup> Da Silva Moore v. Publicis Groupe, 2012 U.S. Dist. LEXIS 23350, at \*3 (S.D.N.Y. Feb. 24, 2012).

<sup>45</sup> *Id.* at \*4.

<sup>46</sup> Da Silva Moore v. Publicis Groupe, 2012 U.S. Dist. LEXIS 58742, at \*3 (S.D.N.Y. April 25, 2012).

<sup>47</sup> *Id.* at \*2. Regarding the plaintiff’s argument on recusal, see Reply Memorandum of Law in Support of Plaintiff’s Motion for Recusal or Disqualification, Da Silva Moore v. Publicis Groupe, Civ. No. 11-cv-1279 (May 10, 2012). As evidence of alleged appearance of partiality, the plaintiffs cite, among other things, certain comments by the magistrate judge (e.g., an alleged comment by Judge Peck that defendants “must have thought they died and went to Heaven” to have him assigned to the case), *id.* at 1, and his participation in educational forums sponsored by various predictive coding vendors. *Id.* at \*6-7. There has been no ruling on the recusal issue, as of May 29, 2012, but in an Order Judge Peck asserted his own impartiality and cautioned the plaintiffs to “rethink their ‘scorched earth approach to this litigation.’” Order at 2, Da Silva Moore v. Publicis Groupe, Civ. No. 11-cv-1279 (April 2, 2012). These colorful allegations and exchanges have contributed to widespread coverage of this case. *See supra* note 11 (citing sampling of blog coverage).

<sup>48</sup> Da Silva Moore v. Publicis Groupe, 2012 U.S. Dist. LEXIS 58742, at \*5-7.

<sup>49</sup> *Id.* at \*5.

<sup>50</sup> *Id.* at \*6-7.

<sup>51</sup> Kleen Prods., LLC v. Packaging Corp. of Am., Civil Case No. 1:10-cv-05711 (N.D. Ill. filed Sept. 9, 2010).

<sup>52</sup> *See* Kleen Prods., LLC v. Packaging Corp. of Am., 775 F. Supp. 2d 1071 (N.D. Ill. 2011) (denying motion to dismiss).

<sup>53</sup> Joint Status Conference Report No. 3 at 3, Kleen Prods., LLC v. Packaging Corp. of Am., Civil Case No. 1:10-cv-05711 (N.D. Ill. May 17, 2012).

<sup>54</sup> Defendant’s Brief on Discovery Issues at 1, Kleen Prods., LLC v. Packaging Corp. of Am., Civil Case No. 1:10-cv-05711 (N.D. Ill. Feb. 6, 2012).

<sup>55</sup> Plaintiff’s Statement of Position with Respect to Disputed Items at 4-5, Kleen Prods., LLC v. Packaging Corp. of Am., Civil Case No. 1:10-cv-05711 (N.D. Ill. Dec. 13, 2011).

---

<sup>56</sup> As occurred in a Virginia state court in April 2012. *See Order Approving the Use of Predictive Coding for Discovery, Global Aerospace, Inc. v. Landow Aviation, L.P.*, Consolidated Case No. CL61040 (Loudon Cnty. April 23, 2012).

<sup>57</sup> “[F]ederal courts have recently emphasized that discovery is a party-driven process requiring efforts at cooperation of the parties between counsel before petitioning the trial court, while endeavoring to control abuses of the system.” PAUL W. GRIMM ET AL., DISCOVERY PROBLEMS AND THEIR SOLUTIONS 292 (2d ed. 2009).

<sup>58</sup> STEPHEN N. SUBRIN & MARGARET Y.K. WOO, LITIGATING IN AMERICA: CIVIL PROCEDURE IN CONTEXT 131 (2006). These authors caution against “the specter of government intrusion,” emphasizing that the process is generally conducted by the parties themselves. *Id.* at 144.

<sup>59</sup> *See* SEVENTH CIR. ELECTRONIC DISCOVERY COMM., PRINCIPLES RELATING TO THE DISCOVERY OF ELECTRONICALLY STORED INFORMATION (Aug. 1, 2010), available at <http://www.discoverypilot.com/> (last visited May 24, 2012).

<sup>60</sup> SEDONA CONF., COOPERATION PROCLAMATION: RESOURCES FOR THE JUDICIARY 4 (Public Comment Version, Aug. 2011), available at <http://www.thesedonaconference.org> (last visited May 24, 2012). For a history of the Sedona Conference, which “convenes working groups of lawyers, judges, academics, and consultants to move the law and the practices of the courts and judges forward in a reasoned way that was both practical and pragmatic,” see Hon. John M. Facciola, *A History of Electronic Discovery*, in MANAGING E-DISCOVERY AND ESI 13, 23 (Michael D. Berman et al. eds., 2011).

<sup>61</sup> STEPHAN LANDSMAN, READINGS ON ADVERSARIAL JUSTICE: THE AMERICAN APPROACH TO ADJUDICATION 34 (1988).

<sup>62</sup> *Id.* at 35 (“Party control has another beneficial effect as well. It affirms human individuality. It mandates respect for the opinions of each party rather than those of his attorney, of the court, or of society at large.”).

<sup>63</sup> *Id.* at 34. *See also generally* Jordan M. Singer, *Proportionality’s Cultural Foundation*, 52 SANTA CLARA L. REV. 145, 162 (2012) (“We are willing to allow wide attorney discretion in conducting pretrial activities because such discretion is the best mechanism we have to promote the ultimate goals (the core values) of a predictable, efficient, and fair resolution on the merits.”).

<sup>64</sup> Fed. R. Civ. P. 26(f)(2). Of course, the parties may not agree on every issue, and in that case, the discovery plan should note their respective positions. The parties are also obligated to discuss at this conference the possibility of settlement, issues connected with preserving potentially relevant ESI, as well as to make disclosures of types and locations of ESI that are required by Fed. R. Civ. P. 26(a)(1). *Id.* The Rules were amended in 2006 specifically to address the special challenges of electronic discovery, including discovery of ESI from sources not readily available and format of production. *See* Ronald J. Hedges, *The ESI Amendments to the Federal Rules of Civil Procedure: A Rule-by-Rule Look*, in MANAGING E-DISCOVERY AND ESI 31, 31-48 (Michael D. Berman et al. eds., 2011); SUBRIN & WOO, *supra* note 58, at 141-43.

<sup>65</sup> Fed. R. Civ. P. 26(f)(3)(B).

<sup>66</sup> Fed. R. Civ. P. 26(f)(3)(D). *See infra* notes 107-16 and accompanying text regarding clawback agreements.

<sup>67</sup> Often a federal magistrate judge, rather than an Article III district court judge, handles pre-trial discovery issues. *See* Stephen S. Gensler, *Judicial Case Management: Caught in the Crossfire*, 60 DUKE L.J. 669, at 694-97 (2010) (discussing role of magistrate judges).

<sup>68</sup> BARBARA J. ROTHSTEIN ET AL., MANAGING DISCOVERY OF ELECTRONIC INFORMATION: A POCKET GUIDE FOR JUDGES 5 (2007). For example, the court may rule on disputes regarding scope of discovery, the allocation of costs, the form of production, the preservation of data, and spoliation. *Id.* at 20-21.

<sup>69</sup> Fed. R. Civ. P. 16(b).

<sup>70</sup> Fed. R. Civ. P. 16(b)(3)(B)(iii).

<sup>71</sup> Fed. R. Civ. P. 16(a)(1).

<sup>72</sup> Fed. R. Civ. P. 16(a)(2).

<sup>73</sup> Fed. R. Civ. P. 16(a)(3). Other listed pretrial functions are “improving the quality of the trial through more thorough preparation,” Fed. R. Civ. P. 16(a)(4), and “facilitating settlement.” Fed. R. Civ. P. 16(a)(5). The functions listed in Fed. R. Civ. P. 16(a) are not exhaustive. *See* Fed. R. Civ. P. 16(a) (indicating that the court may act “for such purposes as . . .”). Common, specific pretrial matters for consideration are listed in Fed. R. Civ. P. 16(c)(2). These include such matters as considering the possibilities for summary judgment, Fed. R. Civ. P. 16(c)(2)(E), and setting deadlines for discovery. Fed. R. Civ. P. 16(c)(2)(F). The listed activities are not exclusive; other activities that go to “facilitating in other ways the just, speedy, and inexpensive disposition of the case” are also appropriate. Fed. R. Civ. P. 16(c)(2)(P).

<sup>74</sup> Fed. R. Civ. P. 1.

<sup>75</sup> *See, e.g.*, Patroski v. Pressley Ridge, 2011 U.S. Dist. LEXIS 133290, \*8 (W.D. Pa. Nov. 17, 2011) (indicating that the court “does not referee discovery bouts between consenting adults”); Murray v. Geithner, 2010 U.S. Dist. LEXIS 33236, at \*2 (S.D.N.Y. March 25, 2010) (“It is not the Court’s task to do [a party’s] . . . job for him by redrafting his manifestly overbroad discovery requests.”).

<sup>76</sup> ROTHSTEIN ET AL., *supra* note 68, at 20.

<sup>77</sup> *See generally* Paul D. Carrington, *Recent Efforts to Change Discovery Rules: Do They Advance the Purposes of Discovery?* 51, 60, in ROSCOE POUND INST., CONTROVERSIES SURROUNDING DISCOVERY AND ITS EFFECT ON THE COURTS

---

(1999), available at [www.roscoepound.org/images/1999ForumReport.pdf](http://www.roscoepound.org/images/1999ForumReport.pdf) (“[C]ase management techniques should not be employed routinely in the absence of evidence that there are abuses to be prevented that cannot be controlled by other means and thus that real benefits can be secured. Judicial involvement in pretrial litigation should be the exception and not the rule.”). Carrington also notes that a move toward case management is a move away from the normal, party-driven process: “The hidden effect of case management is a transfer of power away from individual parties and their lawyers, and [because cases settle rather than going to trial] also from juries or appellate courts who would review decisions on the merits when and if rendered.” *Id.*

<sup>78</sup> Parties engage in gamesmanship when they adhere to form over substance in responding to discovery requests, construing requests narrowly to avoid producing “clearly discoverable material.” *See, e.g., Georgacarakos v. Wiley*, 2011 U.S. Dist. LEXIS 26900, \*16-17 (D. Colo. March 16, 2011) (approving magistrate judge’s *sua sponte* modification of document request, noting that producing party would suffer no prejudice, and that “Constraining a Magistrate Judge’s ability to overlook matters of form and cut to the substantive heart of a discovery dispute would needlessly encourage ‘Gotcha!’-style litigation, where the slightest inaccuracy or deviation from form would allow parties to resist discovery requests”); *Lane v. Page*, 2011 U.S. Dist. LEXIS 21198, \*12 (D. N.M. Feb. 10, 2011) (“The Federal Rules of Civil Procedure embrace a policy of encouraging broad discovery, and exhibit little patience for gamesmanship and attempts to withhold discoverable materials and information. Such gamesmanship produces needless delay, wasting the Court’s and the parties’ time, contrary to the purpose of the Federal Rules of Civil Procedure.”).

<sup>79</sup> *Patroski v. Pressley Ridge*, 2011 U.S. Dist. LEXIS 133290, \*7 (W.D. Pa. Nov. 17, 2011).

<sup>80</sup> *Id.*

<sup>81</sup> *See Jimena v. UBS AG Bank, Inc.*, 2010 U.S. Dist. LEXIS 119393, \*10 (E.D. Cal. Oct. 25, 2010) (“The Rules in place, and the various discretionary sanctions available to the Court to enforce the Rules, are meant to encourage fairness and to avoid obstructionism, gamesmanship, and tactical maneuvering intended to drive up the costs of litigation and unfairly harass the other party.”).

<sup>82</sup> *See SUBRIN & WOO, supra* note 58, at 144.

<sup>83</sup> Fed. R. Civ. P. 26(b)(1).

<sup>84</sup> *Patroski*, 2011 U.S. Dist. LEXIS 133290, at \*8. One scholar encapsulates the judge’s duties: “The judge will not run the discovery program but he will be there to provide guidance and to set limits of time, place, subject matter, and the like as may be appropriate.” Singer, *supra* note 63, at 178. For example, the court can mandate periodic status conferences and specify the form in which electronic documents must be produced. *See Covad Commc’n Co. v. Revonet, Inc.*, 267 F.R.D. 14 (D.D.C. 2010). The court can “limit the frequency or extent of use” of discovery methods. *Cartel Asset Mgmt. v. Ocwen Fin. Corp.*, 2010 U.S. Dist. LEXIS 17857, at \*25 (D. Colo. Feb. 8, 2010). The court can order that questions be answered via interrogatories rather than deposition. *Georgacarakos v. Wiley*, 2011 U.S. Dist. LEXIS 26900, at \*17 (D. Colo. March 16, 2011) (“The Court would be inclined to find that the Magistrate Judge’s decision to prefer one discovery mechanism [interrogatories versus depositions] would not be an abuse of his considerable discretion.”). The court may phase discovery. *Tamburo v. Dworkin*, 2010 U.S. Dist. LEXIS 121510, at \*9 (N.D. Ill. Nov. 17, 2010). *See also generally ROTHSTEIN ET AL., supra* note 68, at 6-9, 16-19 on appropriate tools for managing discovery.

<sup>85</sup> For example, an award of attorneys’ fees may be appropriate if a party has “acted in bad faith, vexatiously, wantonly, or for oppressive reasons.” *See Chambers v. NASCO, Inc.*, 501 U.S. 32, 45-46 (1991) (discussing inherent powers of federal courts and noting that these powers must be exercised with “restraint and discretion”).

<sup>86</sup> *Jimena*, 2010 U.S. Dist. LEXIS 119393, at \*10.

<sup>87</sup> *See Fed. R. Civ. P. 26(g)* advisory committee’s note (1983) (“Concern about discovery abuse has led to widespread recognition that there is a need for more aggressive judicial control and supervision.”); Gensler, *supra* note 67, at 743 (“For nearly thirty years, the Civil Rules have looked to judicial case management as the principal means for controlling excessive cost and delay in civil cases”).

<sup>88</sup> *See SUBRIN & WOO, supra* note 58, at 141-43 (noting that 2006 amendments aimed to ensure that discovery was proportionate and to avoid abuses by expanding role of judges in managing discovery); Gensler, *supra* note 67, at 674-85; Henry S. Noyes, *Good Cause Is Bad Medicine for the New E-Discovery Rules*, 21 HARV. J.L. & TECH. 49, 54-60 (2007) (citing 1983 amendments to Rule 26 allowing the court to limit “overly burdensome” or duplicative discovery and 2006 amendments that require parties to show “good cause” before seeking certain discovery) Noyes criticizes the good cause requirement as overly flexible, inviting widely varying interpretations by judges.

<sup>89</sup> Fed. R. Civ. P. 16(b) advisory committee’s note (2006). *See also BARBARA J. ROTHSTEIN ET AL., supra* note 68, at 15 (Rule 16 “does not authorize the court to require the parties to enter into such an arrangement [clawback agreement], absent their agreement.”).

<sup>90</sup> *See, e.g., Gensler, supra* note 67, at 720-26 (surveying the literature); Jonathan T. Molot, *An Old Judicial Role for a New Litigation Era*, 113 YALE L.J. 27 (2003); Singer, *supra* note 63. *See also JUDICIAL CONFERENCE ADVISORY COMM., supra* note 5, at 22 (noting decades-long debate between those favoring “party-controlled discovery” and those favoring strengthening judges’ tools for managing and/or limiting discovery).

---

<sup>91</sup> See, e.g., Gensler, *supra* note 67, at 723-25.

<sup>92</sup> Molot, *supra* note 90, at 89; Noyes, *supra* note 88, at 78 (“The phenomenon of case management, encouraged by recent amendments to the Rules, hinders the ability of appellate courts to foster uniformity in the law of discovery.”); Singer, *supra* note 63, at 185.

<sup>93</sup> Gensler, *supra* note 67, at 724 (noting “theoretical possibility that trial judges will use their discretion to promote individual substantive agendas”); Arthur R. Miller, *From Conley to Twombly to Iqbal: A Double Play on the Federal Rules of Civil Procedure*, 60 DUKE L.J. 1, 33 (2010) (cautioning that judicial discretion “threatens to become excessive” when it may lead to unpredictability and “reliance on individual predilections”); Noyes, *supra* note 88, at 79 (“Unfettered judicial discretion will further open the door to inconsistent, arbitrary, and biased decision-making.”).

<sup>94</sup> Molot, *supra* note 90, at 89; see also Gensler, *supra* note 67, at 721 (noting this issue).

<sup>95</sup> Noyes, *supra* note 88, at 79. See also Miller, *supra* note 93, at 83 (noting that “unbridled discretion” given judges by pleading standards undermines confidence in justice system and encourages “forum and judge shopping”).

<sup>96</sup> See *supra* notes 33-42 and accompanying text.

<sup>97</sup> See *supra* note 16 and accompanying text.

<sup>98</sup> See *supra* note 34 and accompanying text. Recall, too, that information about particular predictive coding methods may be sparse, leading to the “black box” reputation.

<sup>99</sup> See Fed. R. Civ. P. 1.

<sup>100</sup> Fed. R. Civ. P. 26(g)(1).

<sup>101</sup> Fed. R. Civ. P. 26(g)(1)(A). See also Fed. R. Civ. P. 26(g) advisory committee’s note (1983) (noting that certification requirement is “consistent with the spirit and purposes” of the Rules; to wit, “If primary responsibility is to continue to rest with the litigants, they must be obliged to act responsibly and to avoid abuse.”); Katz, *supra* note 20, at 931-40 (regarding attorneys’ duties in discovery, including ethical duties of competence and diligence).

<sup>102</sup> See *supra* notes 77-82 and accompanying text.

<sup>103</sup> See *supra* Part V.A. and notes 38-42 and accompanying text, citing as possible reasons for slow adoption of predictive coding a lack of technical understanding by lawyers, lack of transparency of the process, concern about accuracy of the results, judicial acceptance, and effect on privilege protection.

<sup>104</sup> See *supra* notes 33-36 and accompanying text.

<sup>105</sup> See, e.g., Custom Hardware Eng’g & Consulting, Inc. v. Dowell, 2012 U.S. Dist. LEXIS 146 (E.D. Mo. Jan. 3, 2012); Osborne v. C.H. Robinson Co., 2011 U.S. Dist. LEXIS 123168 (N.D. Ill. Oct. 25, 2011).

<sup>106</sup> See, e.g., Elizabeth King, *Waving Goodbye to Waiver? Not So Fast: Inadvertent Disclosure, Waiver of the Attorney-Client Privilege, and Federal Rule of Evidence 502*, 32 CAMPBELL L. REV. 467, 467-68 (2010).

<sup>107</sup> On issues of privilege, waiver, and clawback, see generally ABA, THE ATTORNEY-CLIENT PRIVILEGE IN CIVIL LITIGATION: PROTECTING AND DEFENDING CONFIDENTIALITY 206-14 (Vincent S. Walkowiak et al. eds., 2012); John Gallagher, *E-Ethics: The Ethical Dimensions of the Electronic Discovery Amendments to the Federal Rules of Civil Procedure*, 20 GEO. J. LEGAL ETHICS 613, 613-24 (2007).

<sup>108</sup> See Fed. R. Civ. P. 26(f)(3)(D).

<sup>109</sup> Under a clawback arrangement, the parties agree that upon the producing party’s identification of privileged documents mistakenly produced, the receiving party shall return them; the producing party’s privilege has not necessarily been waived. See Fed. R. Civ. P. 26(f) advisory committee’s note (2006). Alternatively, the parties may make a “quick peek” agreement. Under this arrangement, all documents thought to be responsive are produced without any initial privilege review. The receiving party takes a “quick peek” to determine which documents it actually wishes to use; the producing party then screens that smaller set for privilege. *Id.* Clawback or quick peek arrangements may be advantageous because they “can facilitate prompt and economical discovery by reducing delay before the discovering party obtains access to the documents, and by reducing the cost and burden of review by the producing party.” *Id.*

<sup>110</sup> Fed. R. Civ. P. 26(b)(5)(B).

<sup>111</sup> See Fed. R. Evid. 502. This Rule provides that disclosure of attorney-client privileged information does not act as a waiver of the privilege so long as: “(1) the disclosure is inadvertent; (2) the holder of the privilege or protection took reasonable steps to prevent disclosure; and (3) the holder promptly took reasonable steps to rectify the error, including . . . following Federal Rule of Civil Procedure 26(b)(5)(B).” A large body of case law has developed on what sorts of behavior will constitute a waiver. See David D. Cross et al., *Protection From ESI Waiver under FRE 502*, in MANAGING E-DISCOVERY AND ESI 387, 394-407 (Michael D. Berman et al. eds., 2011); King, *supra* note 106, at 470-501. Court-ordered disclosure shall not constitute a waiver, as the disclosure was not voluntary. See, e.g., *Equity Analytics LLC v. Lundin*, 248 F.R.D. 331, 334 (D.D.C. 2008); *Hopson v. Mayor of Balt.*, 232 F.R.D. 228, 241 (D. Md. 2005).

<sup>112</sup> See *supra* notes 24-28 and accompanying text and Figure 2.

<sup>113</sup> Regarding the reasonableness of precautions taken to avoid production of privileged documents, the Advisory Committee specified:

Depending on the circumstances, a party that uses advanced analytical applications and linguistic tools in screening for privilege and work product may be found to have taken “reasonable steps” to prevent

- 
- inadvertent disclosure. ... The Rule does not require the producing party to engage in a post-production review to determine whether any protected communication or information has been produced by mistake. Fed. R. Evid. 502(b) advisory committee's note (2008).
- <sup>114</sup> Cross et al., *supra* note 111, at 408 (“[W]hile an adversary can be barred under FRE from using the document containing that [privileged] information, nothing—not even FRE 502—can undo the potential harm of the privileged information itself having become known to an adversary.”).
- <sup>115</sup> Jessica Wang, Comment, *Nonwaiver Agreements after Federal Rule of Evidence 502: A Glance at Quick-Peek and Clawback Agreements*, 56 UCLA L. REV. 1835, 1846 (2009) (suggesting that knowledge of the privileged information may lead to new witnesses, new topics for exploration, or new arguments for the receiving party, and noting that such illicit use would be difficult to detect or prove); *see also* Hopson, 232 F.R.D. at 241 (noting that rule excluding use as evidence “cannot undo the damage that inadvertent disclosure of privileged information causes, but it can contain it. ‘Confidentiality, once destroyed, is not susceptible of restoration, yet some measures of repair may be accomplished by preventing uses of the evidence against the holder of the privilege.’”).
- <sup>116</sup> *See, e.g.*, Hopson, 232 F.R.D. at 235; ABA, *supra* note 93, at 207.
- <sup>117</sup> Upjohn Co. v. United States, 449 U.S. 383, 389 (1981). *See also* In re Seagate Tech., LLC, 497 F.3d 1360, 1372 (Fed. Cir. 2007).
- <sup>118</sup> See King, *supra* note 106, at 505.
- <sup>119</sup> See *supra*, notes 106-11 and accompanying text.
- <sup>120</sup> *In re Seagate*, 497 F.3d at 1372.
- <sup>121</sup> Parties could agree or a court could order the parties, theoretically, to employ predictive coding but also to do a final pre-production privilege review. Adding such a step, however, would eliminate the promised efficiency of predictive coding. *See supra* notes 26-28 and accompanying text. When adequate privilege protection would compel costly manual review, it would be an unfair burden and contrary to the Rules’ aim of efficiency to order a party opposed to predictive coding to use that method.
- <sup>122</sup> *See supra* notes 83-84 and accompanying text.
- <sup>123</sup> *Id.*
- <sup>124</sup> *See* Fed. R. Civ. P. 26(b)(1) advisory committee’s note (1983) (“Rule 26(b)(1) has been amended to add a sentence to deal with the problem of over-discovery. The objective is to guard against redundant or disproportionate discovery by giving the court authority to reduce the amount of discovery that may be directed to matters that are otherwise proper subjects of inquiry. The new sentence is intended to encourage judges to be more aggressive in identifying and discouraging discovery overuse.”).
- <sup>125</sup> In phased discovery, the most easily retrieved documents are produced and examined first, to aid the parties in determining whether more burdensome discovery would be worthwhile.
- <sup>126</sup> To “limit” is “to curtail or reduce in quantity or extent.” MERRIAM-WEBSTER ONLINE DICTIONARY, <http://www.merriam-webster.com/dictionary/limit?show=1&t=1337905506> (last visited May 24, 2012). *See also* BLACK’S LAW DICTIONARY 834 (5<sup>th</sup> ed. 1979) (to “limit” is “to abridge, confine, restrain, and restrict,” or “to fix the extent of”).
- <sup>127</sup> Robert G. Bone, *Who Decides? A Critical Look at Procedural Discretion*, 28 CARDOZO L. REV. 1961, 1967-74 (2007); *see also* *supra* notes 87-95 and accompanying text.
- <sup>128</sup> *See supra* notes 77-82 and accompanying text.
- <sup>129</sup> Singer, *supra* note 63, at 186.
- <sup>130</sup> *See supra* notes 92-95 and accompanying text. Allegations in *Da Silva Moore* concerning appearance of partiality may give credence to such concerns. *See supra* note 47. Judges must remain above the fray. *See supra* note 63 and accompanying text.
- <sup>131</sup> *See supra* notes 106-21 and accompanying text.
- <sup>132</sup> *See supra* notes 57-63 and accompanying text.
- <sup>133</sup> Note that the Rules do not mandate a perfect search method or the best search method, but rather only a reasonable one. *See supra* note 43 and accompanying text.
- <sup>134</sup> *See, e.g.*, Gary E. Marchant, *Sustainable Energy Technologies: Ten Lessons from the History of Technology Regulation*, 18 WIDENER L.J. 831 (2009) (recounting legislative and executive efforts to foster use of electric cars and compact fluorescent bulbs); Stephen Bates, *Digital Debacle? Lessons from the History of Technical Standards*, 10 MINN. J.L. SCI. & TECH. 441 (2009) (discussing government efforts regarding compact fluorescent light bulbs and switch to digital television).
- <sup>135</sup> Among significant factual differences that limit the analogy include the differing governmental roles of judges as compared with legislators or administrators; the types of harm that improved technology is thought to prevent (less accurate or costlier discovery as compared with environmental harm); and the type of “consumer” (attorneys conducting discovery in large-scale litigation as compared with individuals buying light bulbs or cars).
- <sup>136</sup> *E.g.*, Marchant, *supra* note 134, at 834-35.
- <sup>137</sup> *Id.* at 835.

---

<sup>138</sup> *Id.* at 836 (“The historical record is that governments . . . have a relatively poor record in picking which future technologies will best succeed in achieving a particular objective.”); *see also* Bates, *supra* note 134, at 448 (“The ‘best’ standard can be difficult to identify.”).

<sup>139</sup> Marchant, *supra* note 134, at 845.

<sup>140</sup> *Id.* at 855.