

**2009 Annual Meeting**  
**The Association of Life Insurance Counsel**  
May 3, 2009 – May 5, 2009

**E-Discovery:  
Addressing The Information Logjam  
Without Getting Rolled**

**Donald A. Murday, Esq.**  
Chittenden, Murday & Novotny, LLC  
303 W. Madison Street, Suite 1400  
Chicago, Illinois 60606  
(312) 281-3600  
dmurday@cmn-law.com

## TABLE OF CONTENTS

I.	Introduction .....	1
II.	Federal Rules of Civil Procedure Related to E-Discovery .....	1
	A. Rule 26.....	1
	1. Rule 26(a)(1)(A)(ii): Initial Disclosures.....	1
	2. Rule 26(b)(2)(B)–(C): Accessibility of Data.....	2
	3. Rule 26(b)(5)(B): The Claw-Back Provision .....	2
	4. Rule 502 of the Federal Rules of Evidence .....	3
	a. Subject-Matter Waiver .....	3
	b. Requirements for Avoiding Waiver of Privilege in Context of Inadvertent Disclosure.....	3
	5. Rule 26(c): Cost-Shifting.....	4
	6. Rule 26(f)(2): E-Discovery at the Initial Attorneys' Conference.....	5
	B. Rule 16(b)(3)(B): Discussing E-Discovery at the Outset of Litigation.....	5
	C. Rule 33(d): Interrogatories and the Production of ESI.....	5
	D. Rule 34: Document Requests and the Production of ESI.....	5
	E. Rule 37(e): A "Safe Harbor" for Routine Destruction.....	6
	F. Rule 45: Subpoenas for ESI.....	6
	G. Rule 53: The Use of Special Masters and E-Discovery .....	7
III.	Issues Confronting Corporate Law Departments Regarding the Management and Production of ESI in Discovery .....	7
	A. New Federal Rule of Evidence 502 Protects Parties From the Waiver of Privileged Information Due to Inadvertent Disclosure.....	7
	1. Reasonable Steps to Prevent Disclosure and to Rectify the Error.....	7
	2. Rule 502 and the Relationship Between Federal and State Courts .....	8
	B. Limiting Discovery Costs at the Outset Through Rule 26(f) Conferences .....	10
	C. Establishing a Litigation or Legal Hold .....	11
	1. The Trouble With Using Keyword Searches to Identify ESI That Must Be Preserved .....	13

2.	It May No Longer Be Reasonable to Rely on Keyword Searches Alone to Identify ESI That Must Be Preserved.....	15
3.	If Keyword Searches Are Not Enough, What Search Methodologies Are Reasonable?.....	18
	a. Fuzzy Search Models.....	18
	b. Concept Search Models.....	18
	i. Latent Semantic Analysis.....	18
	ii. Text Clustering.....	19
	iii. Bayesian Classifier.....	19
D.	The Need For and Selection of Third-Party E-Discovery Vendors.....	20
	1. Define the Services Required.....	21
	a. Categories of Vendor Services.....	21
	b. Requesting Information.....	21
	2. Examining Potential Vendors.....	23
	a. Qualifications and Reputation.....	23
	b. Security.....	23
	c. Conflicts.....	23
	d. Pricing Models.....	24
	3. Soliciting Proposals and Making the Selection.....	24
E.	Has Ms. Zubulake Turned Your Retention Program Into a De Facto "Keep Everything Forever" Program?.....	26
	1. Best Practices for Document Retention Policy Creation and Enforcement ....	27
	2. A Proposed Plan for Preventing Litigation Holds from Eclipsing a Company's Document Retention Policy.....	29
	3. There Are No Easy Answers for Dealing with a Backlog of Documents and ESI.....	31
IV.	Conclusion.....	32

## **I. Introduction**

The years 2005 through early 2008 saw a “revolution” in e-discovery during which courts, the legislature, and litigants worked hard to establish the ground rules for e-discovery. In 2005 and 2006, the focus was on the duty to preserve and the consequences of spoliation of electronically stored information (“ESI”), as well as the management of ESI-related discovery under newly-enacted Federal Rules of Civil Procedure. In 2007 and early 2008, courts and commentators helped establish best practices for the component parts that make up the management of an e-discovery project. The development of e-discovery has now shifted from “revolution” to “evolution.” The focus of recent court decisions, legislative activity, and commentary has been to build on the ground rules for e-discovery. For instance, in September 2008, new Federal Rule of Evidence 502 went into effect, allowing parties to agree that the inadvertent disclosure of privileged materials produced under an agreed discovery protocol would not amount to a waiver of a privilege. Importantly, non-waiver orders under Rule 502 are binding in other federal and state court actions. One of the new Rule’s primary purposes is to reduce the often burdensome discovery costs associated with conducting a privilege review of extensive ESI prior to production in discovery. Also in 2008, federal courts began to examine the methodologies by which parties search for relevant ESI to preserve and produce. Those decisions are significant because they are the first steps in the evolution of conduct standards that will hopefully help make the search, retrieval, production, and management of ESI simpler, more accurate, and more cost effective. The result is that parties now have a clearer idea of how to manage the challenges and mitigate the risks associated with e-discovery.

This paper first briefly discusses the Federal Rules of Civil Procedure related to e-discovery in order to provide a baseline understanding of the federal e-discovery framework. The paper also reviews the continuing developments regarding e-discovery and anticipates their impact going forward. Finally, the paper offers practical advice concerning litigation hold processes that limit interference with a company’s document retention program, the selection of third-party vendors, and the ways to use the required scheduling conferences to limit discovery costs.

## **II. Federal Rules Of Civil Procedure Related To E-Discovery**

Several revisions to the Federal Rules of Civil Procedure in 2006 and 2007, and the enactment of Rule 502 of the Federal Rules of Evidence in 2008 concern e-discovery issues. A familiarity with these rules is essential to an understanding of the various issues companies face when managing the retention and production of what can often become a “logjam” of ESI.

### **A. Rule 26**

#### **1. Rule 26(a)(1)(A)(ii): Initial Disclosures**

Rule 26(a)(1) governs parties’ initial disclosures of relevant information. Under Rule 26(a)(1)(A)(ii), required disclosures include among the types of information a litigant must provide to its opponent “a copy—or a description by category and location —of all documents,

electronically stored information and tangible things that the disclosing party has in its possession, custody or control and may use to support its claims or defenses, unless the use would be solely for impeachment.”<sup>1</sup> “The term ‘electronically stored information’ has the same broad meaning ... as in Rule 34(a).”<sup>2</sup>

## 2. Rule 26(b)(2)(B)–(C): Accessibility Of Data

Rule 26(b)(2)(B) provides:

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. On motion to compel discovery or for a protective order, the party from whom discovery is sought must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.<sup>3</sup>

Rule 26(b)(2)(B), however, “has been criticized for giving the parties to a lawsuit too much leeway in determining whether data is accessible.”<sup>4</sup>

Further, Rule 26(b)(2)(C) provides:

On motion or on its own, the court must limit the frequency or extent of discovery otherwise allowed by these rules or by local rule if it determines that:

- (i) the discovery sought is unreasonably cumulative or duplicative, or can be obtained from some other source that is more convenient, less burdensome, or less expensive;
- (ii) the party seeking discovery has had ample opportunity to obtain the information by discovery in the action; or
- (iii) 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 importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.<sup>5</sup>

## 3. Rule 26(b)(5)(B): The Claw-Back Provision

Rule 26(b)(5) provides a method by which counsel can retrieve inadvertently produced material that is privileged or work-product. Rule 26(b)(5)(B) describes what has been called the ‘claw-back’ agreement:

[i]f information produced in discovery is subject to a claim of privilege or of protection as trial-preparation material, the party making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information to the court under seal for a determination of the claim. The producing party must preserve the information until the claim is resolved.<sup>6</sup>

“Quick-peek” agreements are similar to “claw-back” agreements. Because of the potential high cost of the initial review for privileged materials, “[p]arties may attempt to minimize ... costs and delays by agreeing to protocols that minimize the risk of waiver. They may agree that the responding party will provide certain requested materials for initial examination without waiving any privilege or protection – sometimes known as a ‘quick peek.’”<sup>7</sup> Before the enactment of Rule 502 of the Federal Rules of Evidence (see *infra*), “claw-back” and “quick-peek” agreements were risky because they “did not bind those who were nonparties, thus providing no protection against the risk that a third party could seek to obtain and use the documents in another proceeding.”<sup>8</sup>

#### **4. Rule 502 Of The Federal Rules Of Evidence**

Rule 502 of the Federal Rules of Evidence<sup>9</sup>, which became effective on September 19, 2008, is related to Rule 26(b)(5)(B) and addresses the consequences of inadvertently disclosing privileged material. The most significant aspect of the rule is that it allows Federal courts to issue orders that a privilege has not been waived and makes those orders binding on non-parties, as well as on other Federal and State courts. The operative sections of Rule 502 are summarized below, but for a thorough discussion of the new Rule, see Section II. A. of this paper.

##### **a. Subject-Matter Waiver**

Federal Rule of Evidence 502(a) provides that: “When the disclosure is made in a Federal proceeding or to a Federal office or agency and waives the attorney-client privilege or work-product protection, the waiver extends to an undisclosed communication or information in a Federal or State proceeding only if: (1) the waiver is intentional; (2) the disclosed and undisclosed communications or information concern the same subject matter; and (3) they ought in fairness to be considered together.”<sup>10</sup> “In effect, an inadvertent disclosure, even if it constitutes a waiver, will act as a waiver only as to the materials disclosed, not to other materials regarding the same subject matter.”<sup>11</sup>

##### **b. Requirements For Avoiding Waiver Of Privilege In Context Of Inadvertent Disclosure**

Rule 502(b) addresses the potential effect of an inadvertent disclosure of privileged or work-product material on a claim of privilege: “(b) Inadvertent disclosure.- - When made in a

Federal proceeding or to a Federal office or agency, the disclosure does not operate as a waiver in a Federal or State proceeding if: (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 (if applicable) following Federal Rule of Civil Procedure 26(b)(5)(B).<sup>12</sup> The rule defines “attorney-client privilege” as “the protection that applicable law provides for confidential attorney-client communications,” and “work-product protection” as “the protection that applicable law provides for tangible material (or its intangible equivalent) prepared in anticipation of litigation or for trial.”<sup>13</sup> Rule 502 “applies to previously filed cases, in the discretion of the trial judge, and to all cases filed on and after September 19, 2008, without exception.”<sup>14</sup>

### 5. Rule 26(c): Cost-Shifting

Retrieving and producing ESI raises cost concerns. While the presumption “is that the responding party must bear the expense of complying with discovery requests ... it may invoke the district court’s discretion under Rule 26(c) to grant orders protecting it from undue burden or expense in doing so, including orders conditioning discovery on the requesting party’s payment of the costs of discovery.”<sup>15</sup> Rule 26(c)(1) provides in pertinent part: “[t]he court may, for good cause, issue an order to protect a party or person from annoyance, embarrassment, oppression, or undue burden or expense ....”<sup>16</sup>

The United States District Court for the Southern District of New York in *Rowe Entertainment, Inc. v. The William Morris Agency*<sup>17</sup> noted “courts have adopted a balancing approach,” and apply the following eight-factor cost-shifting test:

- (1) the specificity of the discovery requests; (2) the likelihood of discovering critical information; (3) the availability of such information from other sources; (4) the purposes for which the responding party maintains the requested data[;] (5) the relative benefit to the parties of obtaining the information; (6) the total cost associated with production; (7) the relative ability of each party to control costs and its incentive to do so; and (8) the resources available to each party.<sup>18</sup>

In a widely followed case, the same court in *Zubulake v. UBS Warburg LLC*<sup>19</sup> concluded courts should consider shifting the costs of production to requesting parties when the requested information is “inaccessible” - when it must be restored or reconstructed before it can be used or reviewed.<sup>20</sup> Moreover, the *Zubulake* court noted:

[i]f information is inaccessible, the court must weigh seven factors to determine whether it is appropriate to shift the costs of producing the requested information to the requesting party. Those factors are:

- a. The extent to which the request is specifically tailored to discover relevant information;
- b. The availability of such information from other sources;

- c. The total cost of production, compared to the amount in controversy;
- d. The total cost of production, compared to the resources available to each party;
- e. The relative ability of each party to control costs and its incentive to do so;
- f. The importance of the issues at stake in the litigation; and
- g. The relative benefits to the parties of obtaining the information.<sup>21</sup>

#### 6. **Rule 26(f)(2): E-Discovery At The Initial Attorneys' Conference**

Rule 26(f)(2) expressly requires parties to “discuss any issues relating to preserving discoverable information” at the initial conference.<sup>22</sup> Specifically, the Rule requires the parties to “develop a proposed discovery plan”<sup>23</sup> that states “the parties’ views and proposals on ... any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced ... [and] any issues about claims of privilege or of protection as trial-preparation materials, including—if the parties agree on a procedure to assert such claims after production—whether to ask the court to include their agreement in an order ...”<sup>24</sup> “The Advisory Committee Note expressly discourages courts from entering blanket preservation orders and suggests that any preservation order be narrowly tailored.”<sup>25</sup> For a more detailed discussion, see Section III.A., *infra*.

#### B. **Rule 16(b)(3)(B): Discussing E-Discovery At The Outset Of Litigation**

Scheduling orders courts enter often follow the parties’ initial conference scheduling agreements. Accordingly, Rule 16(b)(3)(B) provides, in pertinent part, “[t]he scheduling order may . . . provide for disclosure or discovery of electronically stored information . . . [and] include any agreements the parties reach for asserting claims of privilege or of protection as trial-preparation material after information is produced ....”<sup>26</sup>

#### C. **Rule 33(d): Interrogatories And The Production Of ESI**

Rule 33(d) provides in pertinent part: “[i]f the answer to an interrogatory may be determined by examining, auditing, compiling, abstracting, or summarizing a party’s business records (including electronically stored information), and if the burden of deriving or ascertaining the answer will be substantially the same for either party, the responding party may answer by ... specifying the records that must be reviewed ....”<sup>27</sup> Thus, under the Rule, a party has the option of specifying ESI in response to a written interrogatory.

#### D. **Rule 34: Document Requests And The Production Of ESI**

Similarly, Rule 34(a)(1)(A) provides a party may request “any designated documents or electronically stored information—including writings, drawings, graphs, charts, photographs,

sound recordings, images, and other data or data compilations – stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form ...”<sup>28</sup> Rule 34(a)(1) also allows parties to “inspect, copy, test or sample” the documents, ESI, and tangible things covered by that Rule.<sup>29</sup>

Further, Rule 34(b)(1)(C) allows the requesting party to “specify the form or forms in which electronically stored information is to be produced.”<sup>30</sup> “The response may state an objection to a requested form for producing electronically stored information. If the responding party objects to a requested form—or if no form was specified in the request—the party must state the form or forms it intends to use.”<sup>31</sup> Also, “[i]f a request does not specify a form for producing electronically stored information, a party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms ....”<sup>32</sup>

#### **E. Rule 37(e): A “Safe Harbor” For Routine Destruction**

Rule 37(e) provides: “[a]bsent exceptional circumstances, a court may not impose sanctions under these rules on a party for failing to provide electronically stored information lost as a result of the routine, good-faith operation of an electronic information system.”<sup>33</sup> “Good faith is generally understood to be the absence of bad faith, so if a spoliating party can show that its actions were not in bad faith, it will have met the state of mind standard required by Rule 37(e).”<sup>34</sup>

This rule has been described as providing a “safe harbor” for litigants, who destroy documents in the routine course of electronic data management.<sup>35</sup> Rule 37(e) has also been described as providing “a very shallow harbor.”<sup>36</sup> Further, “courts will likely assess the intent of a producing party (culpable state of mind) that is unable to produce relevant information because it was not maintained, as well as the prejudice to the requesting party from the inability to obtain such data.”<sup>37</sup>

#### **F. Rule 45: Subpoenas For ESI**

Rule 45(a)(1)(C) provides in pertinent part: “[a] command to produce documents, electronically stored information, or tangible things or to permit the inspection of premises may be included in a subpoena commanding attendance at a deposition, hearing, or trial, or may be set out in a separate subpoena. A subpoena may specify the form or forms in which electronically stored information is to be produced.”<sup>38</sup> Also, Rule 45(a)(1)(D) provides, “[a] command in a subpoena to produce documents, electronically stored information, or tangible things requires the responding party to permit inspection, copying, testing, or sampling of the materials.”<sup>39</sup>

## **G. Rule 53: The Use Of Special Masters And E-Discovery**

Rule 53(a) provides that:

[u]nless a statute provides otherwise, a court may appoint a master only to: (A) perform duties consented to by the parties; (B) hold trial proceedings and make or recommend findings of fact on issues to be decided without a jury if appointment is warranted by: (i) some exceptional condition; or (ii) the need to perform an accounting or resolve a difficult computation of damages; or (C) address pretrial and posttrial matters that cannot be effectively and timely addressed by an available district judge or magistrate judge of the district.<sup>40</sup>

Also, “[u]nless the appointing order directs otherwise, a master may: (A) regulate all proceedings; (B) take all appropriate measures to perform the assigned duties fairly and efficiently; and (C) if conducting an evidentiary hearing, exercise the appointing court’s power to compel, take, and record evidence.”<sup>41</sup> “Courts are appointing special masters to address electronic discovery issues with increasing frequency, although the number of reported appointments is still relatively small.”<sup>42</sup> Special masters serve various roles regarding electronic discovery: “(1) facilitating the electronic discovery process; (2) monitoring discovery compliance related to ESI; (3) adjudicating legal disputes related to ESI; and (4) adjudicating technical disputes and assisting with compliance on technical matters, such as conducting computer/system inspections.”<sup>43</sup>

## **III. Issues Confronting Corporate Law Departments Regarding The Management And Production Of ESI In Discovery**

### **A. New Federal Rule Of Evidence 502 Protects Parties From The Waiver Of Privileged Information Due To Inadvertent Disclosure**

Before Rule 502 was enacted, Federal courts generally protected parties from waiving a privilege regarding inadvertently disclosed material, “unless the disclosing party was negligent in producing the information or failed to take reasonable steps seeking its return.”<sup>44</sup> Some courts, however, held that any inadvertent disclosure of privileged material constituted a waiver of the privilege.<sup>45</sup> Others required that the disclosure be intentional in order to result in a waiver.<sup>46</sup> As stated in the Explanatory Note, Rule 502(b) “opts for the middle ground: inadvertent disclosure of protected communications or information in connection with a federal proceeding or to a federal office or agency does not constitute a waiver if the holder took reasonable steps to prevent disclosure and also promptly took reasonable steps to rectify the error. This position is in accord with the majority view on whether inadvertent disclosure is a waiver.”<sup>47</sup>

#### **1. Reasonable Steps To Prevent Disclosure And To Rectify The Error**

Counsel must not assume that all inadvertent disclosures of material will be exempt from a waiver of privilege. Satisfying the court that a party has taken reasonable steps both to avoid

the disclosure and to correct the mistake is crucial. Rule 502(b) does not define what the “reasonable steps” are, but “[b]ecause the reasonableness test adopted in Rule 502 is taken from the majority rule developed over many years, many courts have already addressed whether a party took reasonable steps to avoid inadvertent waivers.”<sup>48</sup> The Advisory Committee Notes provide only a limited list of reasonableness factors, such as “the number of documents to be reviewed and the time constraints for production,” using “software applications and linguistic tools in screening for privilege,” and “an efficient system of records management.”<sup>49</sup>

In any event, some common-sense approaches during the search for and production of materials will help to avoid inadvertent production of privileged materials in the first place. An obvious reasonable step to take is to carefully review *all* documents and materials before producing them. In one case, for example, a party produced a privileged document from a database it thought contained only non-privileged materials. Despite an applicable “claw-back” provision, the court held the party waived the privilege because it did not review the database for privilege before producing the material.<sup>50</sup> The most cautious (and perhaps most costly) approach is to make sure attorneys review all materials for privilege. “Even relatively tolerant courts have demonstrated their disapproval of nonlawyers handling privileged documents.”<sup>51</sup> Further, as one commentator asserted, especially when outside vendors handle the copying of materials, “there must be a final review of documents before they are produced to opposing counsel ... consist[ing] of a face check of each document to make sure that counsel is not producing privileged material.”<sup>52</sup> In addition, “all privileged documents [including electronically produced documents] should be clearly labeled and adequately separated from nonprivileged responsive documents.”<sup>53</sup>

The type of search conducted and the legibility of the material are other important factors. Parties must also pay close attention to the effectiveness of their keyword searches for privileged material.<sup>54</sup>

An additional factor to consider is the time within which a party seeks to rectify the error after discovering it. For example, one court agreed the privilege was not waived where attorneys took steps to correct the inadvertent disclosure “immediately” after realizing the disclosure.<sup>55</sup> Another court found that the plaintiff’s delay in reacting after learning of the inadvertent disclosure, including taking two weeks to determine how the disclosure occurred, was not reasonable.<sup>56</sup>

The parties may also define the “reasonable steps” during the Rule 26(f) conference.<sup>57</sup> For example, the parties could agree to specific time periods within which they could seek to recover inadvertently disclosed materials. In addition to removing as much ambiguity as possible, doing so would provide a “checklist” of reasonable steps that could be circulated to the individuals involved in the search for and production of materials.

## 2. Rule 502 And The Relationship Between Federal And State Courts

The orders entered under Rule 502 bind non-parties and other courts, including State courts. Some commentators have noted that Rule 502 may face constitutional challenges with

respect to this binding effect on State courts.<sup>58</sup> Accordingly, until the constitutionality of the binding effect on State courts has been resolved, parties should still do as much as possible to avoid the inadvertent production of privileged material in the first place.

The interplay among Rule 502 and other Federal Rules of Evidence and the Constitution reveals a potential constitutional pitfall. Specifically, Rule 502(c) provides that:

[w]hen the disclosure is made in a State proceeding and is not the subject of a State-court order concerning waiver, the disclosure does not operate as a waiver in a Federal proceeding if the disclosure:

- (1) would not be a waiver under this rule if it had been made in a Federal proceeding; or
- (2) is not a waiver under the law of the State where the disclosure occurred.<sup>59</sup>

“In effect, the new rule [502] requires the federal court to apply the law that is most protective of the attorney-client privilege and work-product protection.”<sup>60</sup> At the same time, however, Section (f) provides that “[n]otwithstanding Rules 101<sup>61</sup> and 1101<sup>62</sup>, this rule applies to State proceedings and to Federal court-annexed and Federal court-mandated arbitration proceedings, in the circumstances set out in the rule. And notwithstanding Rule 501<sup>63</sup>, this rule applies even if State law provides the rule of decision.”<sup>64</sup> Rule 502(e) provides that “[a]n agreement on the effect of disclosure in a Federal proceeding is binding only on the parties to the agreement, *unless it is incorporated into a court order.*”<sup>65</sup> Section (d) of the rule provides, however, that “[a] Federal court may order that the privilege or protection is not waived by disclosure connected with the litigation pending before the court—in which event the disclosure is also not a waiver in any other Federal or State proceeding.”<sup>66</sup> This has been described as “the ‘most critical piece’ of the legislation,” because the contemplated court orders “bind[] all other courts, as well as nonparties.”<sup>67</sup>

Accordingly, as one commentator noted:

There may be serious constitutional questions as to the ability of Congress to enact such a law merely by asserting that the interest in the federal objective of limiting the costs of production requires such a rule. In *Erie Railroad Co. v. Tompkins*, Justice Brandeis wrote, ‘Congress has no power to declare substantive rules of common law applicable in a state .... And no clause in the Constitution purports to confer such a power upon the federal courts.’ ... To the extent that Rule 502 may overrule the *Erie* doctrine by encroaching on substantive privilege law that has traditionally been left to the states, the rule as drafted poses difficult constitutional questions ....<sup>68</sup>

Further, “[t]he legitimacy of a rule that would bind states or federal agencies in a subsequent litigation may ultimately depend on Congress’s power under the Commerce Clause.”<sup>69</sup> This paper raises the constitutional issue to illustrate that even with the enactment of Rule 502, parties

should not become complacent and assume that inadvertent disclosures of privileged materials will not have any damaging effects.

### **B. Limiting Discovery Costs At The Outset Through Rule 26(f) Conferences**

Fully addressing the preservation and production of ESI in a Rule 26(f) conference has a direct bearing on e-discovery costs. Parties understand how expensive it is to preserve, produce, and review ESI. Defendants know that broad litigation hold notices can be disruptive to their businesses and expensive to maintain.<sup>70</sup> Plaintiffs understand that if they propound (or are forced to respond to) overly broad discovery for ESI, they are likely to be cursed with an ESI production that is difficult to manage and expensive to review.<sup>71</sup> The Rule 26(f) conference gives parties an opportunity to limit the scope of their preservation and production obligations and, thereby limit their exposure to exorbitant discovery costs.

Rule 26(f)(2) directs the parties to “discuss any issues about preserving discoverable information” and to “develop a proposed discovery plan.”<sup>72</sup> Parties should use the opportunity to agree to the scope of their preservation obligations. Depending on the case, parties might agree to limitations on: (a) date ranges; (b) the systems containing ESI to which the preservation obligation will or will not apply; and (c) search terms and/or methodologies to be used in identifying ESI that must be preserved.<sup>73</sup>

Agreeing to date ranges on the parties' preservation obligations is an important part of limiting the costs related to ESI. A party's preservation obligation is typically broader than its production obligation. Thus, limiting the scope of the duty to preserve will also limit the scope of the party's obligation to search for and produce ESI in discovery. Limiting the preservation obligation to ESI that falls within a defined range of dates can substantially reduce preservation and collection costs because the party will only have to identify and collect the ESI one time.<sup>74</sup> Due to the nature of the dispute or the availability of information, parties are typically able to agree to limit the preservation obligation based on a so-called “front-end” date, *i.e.*, the date after which all relevant ESI must be preserved, but an agreed upon “back-end” date can be more elusive. As one author noted, the inability to agree on a back-end cut-off date “places a great burden on a corporation because its employees must preserve documents as they are created, sent or received during the pendency of the litigation.”<sup>75</sup> In the event a back-end date cannot be negotiated, the parties should at least attempt to otherwise narrow the scope of the documents that must be preserved going forward.<sup>76</sup>

Another way to limit e-discovery costs and exposure to spoliation is to agree that the parties are only required to preserve, search, and produce information from specified network systems.<sup>77</sup> For example, parties might agree to limit their preservation obligations to ESI located on their e-mail systems, core office document systems, and databases that store certain types of ESI relevant to the parties' particular dispute.<sup>78</sup> If a party hopes to limit its obligations in this way, it has to be prepared to exchange with the other parties the technical details of its information and network systems so all parties can evaluate a proposed limit on the scope of their preservation obligations.<sup>79</sup>

Obviously, parties need not and cannot agree on all of the issues mentioned above at one meeting. Given the number and complexity of the topics the parties are called on to address in their Rule 26(f) conference, they will likely have to meet more than once (and possibly several times) before they are able to agree on ways to define their preservation obligations and to search for and limit discovery of ESI. Reaching an early consensus on those issues, however, “has the potential to minimize the overall time, cost, and resources spent on [ESI searches], as well as minimizing the risk of collateral litigation” challenging the reasonableness of the scope of a parties’ litigation hold and its searches for ESI.<sup>80</sup>

### C. Establishing A Litigation Or Legal Hold

Virtually all e-discovery risk is concentrated in the litigation-hold process.<sup>81</sup> Identifying the individuals with knowledge of the dispute, conducting reasonable searches of a party’s ESI and paper documents to identify materials that must be preserved, and communicating the hold requirements to others in an organization are critical to ensuring a party is protected against liability or even sanctions for spoliation. The failure to sufficiently create, apply, and enforce a litigation hold can lead to claims of spoliation of evidence and severe sanctions.

“Spoliation is the ‘destruction or significant alteration of evidence, or the failure to preserve property for another’s use as evidence in pending or reasonably foreseeable litigation.’”<sup>82</sup> Potential sanctions for spoliation of evidence include the dismissal of a claim or the granting of summary judgment in favor of the party who is prejudiced, an adverse inference jury instruction, fines, and attorney’s fees and costs.<sup>83</sup> The following cases illustrate the consequences to parties who fail to preserve ESI. The most serious sanctions imposed therein could have been avoided with a more established, reasonable litigation hold protocol, diligent compliance, and accurate representations to opposing counsel and the court as to what information was accessible and inaccessible.

Certain notable cases involving sanctions endure. For example, in *Zubulake v. UBS Warburg LLC*,<sup>84</sup> one in a line of landmark electronic discovery decisions, the court found that the defendant negligently failed to preserve relevant information on certain backup tapes and recklessly failed to preserve relevant information on others. The defendant advised, however, that it discovered new evidence that contained some of the information that it negligently and/or recklessly destroyed. The court held that an adverse inference against the defendant in connection with the destroyed evidence was not appropriate unless the plaintiff could “demonstrate not only that [the defendant] destroyed relevant evidence as that term is ordinarily understood, but also that the destroyed evidence would have been favorable to [the plaintiff].” Absent such a showing, the court held that an adverse inference would *only* be appropriate if the defendant’s destruction of evidence were *willful*.<sup>85</sup> The court, however, required the defendant to pay for the plaintiff’s re-depositions of individuals in connection with the destruction of evidence and the issues raised by the newly discovered information.

The *Zubulake* discovery disputes continued after the plaintiff re-deposed certain witnesses. The depositions revealed that the defendant deleted more e-mails than previously believed and that some of the e-mails thought to have been destroyed were, in fact, preserved on

the defendant's active servers but never produced.<sup>86</sup> Ruling on the plaintiff's motion for sanctions, the court found that the defendant's attorneys failed to effectively communicate the litigation hold to all "key players" and failed to ascertain each of their document management habits.<sup>87</sup> The court also found that some of the key players defied the retention instructions they received. The court concluded that the defendant acted willfully in destroying potentially relevant information, which resulted either in the absence of such information or its tardy production.<sup>88</sup> Further, because the spoliation was willful, the lost information was presumed relevant.<sup>89</sup> Accordingly, the court ordered the defendant to pay the costs of re-deposing the personnel who revealed the newly-discovered e-mails and to pay all costs and attorneys' fees associated with the plaintiff's motion for sanctions

More recently, in *Keithley v. Homestore.com, Inc.*,<sup>90</sup> which involved what the court described as some of the most egregious discovery misconduct the District Court for the Northern District of California has seen, the magistrate judge granted, in part, the plaintiffs' request for sanctions for misconduct due to the defendants' "belated production of evidence that it had previously stated was either nonexistent or destroyed."<sup>91</sup> In her order, the magistrate judge noted that late production, not just frustrated production due to document destruction, is sanctionable.<sup>92</sup> The defendants, which included an Internet real estate search company, lacked a written litigation hold policy at all relevant times and even at the time of the court's sanction order.<sup>93</sup> The magistrate judge recommended to the district court an adverse inference instruction, awarded a monetary sanction of approximately \$185,000, and suggested that she might entertain additional sanctions in excess of \$800,000 for forensic analysis costs.<sup>94</sup>

In *Arista Records LLC v. Usenet.com, Inc.*,<sup>95</sup> the court granted, in part, plaintiff's motion for sanctions for allowing or causing usage data, digital music files, and website promotional materials to be destroyed or withheld in a copyright infringement case. The court held that the defendants had a duty to preserve all evidence at issue in the motion.<sup>96</sup> Even if the defendants' actions with regard to their computer servers did not render certain data unusable, defendants' counsel apparently only agreed to produce a "snapshot" of the requested data based on a ridiculously narrow and naïve keyword search for the words "mp3" or "sound."<sup>97</sup> The court ordered an adverse inference instruction, a sanction less drastic than the plaintiffs' proposed findings of fact but still "serious," as well as attorneys' fees and costs incurred in connection with the motion.<sup>98</sup>

Another case known for its harsh sanctions is *Coleman (Parent) Holdings, Inc. v. Morgan Stanley & Co., Inc.*<sup>99</sup> Unlike the foregoing cases, the sanctions the court imposed in that case were not due to a failure to preserve relevant ESI. Rather, the sanctions were due to the defendant's failure to produce information requested in discovery and required by court order. Despite the lack of a spoliation issue, the decision is significant because it further highlights the nature of the sanctions to which parties can be exposed when it comes to ESI.

The plaintiff, Coleman (Parent) Holdings, Inc., sued Morgan Stanley & Co., Inc. for fraud and sought over \$485 million in damages. To establish Morgan Stanley's knowledge of the fraud, the plaintiff requested Morgan Stanley review certain backup tapes and produce

certain e-mails stored on those tapes. The court ordered Morgan Stanley to comply with the request and further ordered it to certify that its production complied with the court's order.

Morgan Stanley produced 1,300 e-mails and certified that its production was complete. Morgan Stanley knew, however, that its certification was false. In fact, Morgan Stanley was aware of an additional 2,161 backup tapes that had not been reviewed and from which no e-mails were produced. Morgan Stanley waited five months to inform the court and opposing counsel that its earlier certification was not accurate and, the next day, produced an additional 8,000 pages of e-mails to the plaintiff. Two months later, Morgan Stanley notified the court and the plaintiff that it found an additional 169 backup tapes. One month later, on the eve of the plaintiff's motion to instruct the jury that Morgan Stanley's conduct could give rise to the inference that lost or destroyed e-mails were harmful to Morgan Stanley, Morgan Stanley informed the court that: (i) it neglected to produce some of the attachments to the e-mails it earlier produced; (ii) it located another 200 backup tapes that it believed were relevant but which had not yet been reviewed; (iii) a flaw in Morgan Stanley's earlier search for e-mail erroneously omitted at least 7,000 additional e-mail messages that fell within the scope of the court's production order; and (iv) it found 73 bankers boxes of backup tapes but did not know how many of the tapes fell within the scope of the court's production order. The court found that Morgan Stanley lied about its efforts to retrieve electronic documents and "deliberately and contumaciously violated numerous discovery orders." The court also found that Morgan Stanley wrote over e-mails, "contrary to its legal obligation to maintain them in readily accessible form for two years and with knowledge that legal action was threatened."

The court ordered that an extensive statement of conclusive facts be read to the jury describing Sunbeam and Morgan Stanley's fraudulent scheme. In part, the statement reads that "Morgan Stanley conspired with [Sunbeam] to conceal the truth about Sunbeam's financial performance and business operations," and that "Morgan Stanley committed overt acts in furtherance of the conspiracy." The court further ordered that the jury be instructed that those facts were deemed established for all purposes in the action, and that the jury could consider those facts when determining whether Morgan Stanley "sought to conceal its offensive conduct when determining whether an award of punitive damages is appropriate." Finally, the court switched the burden of proof in the case. Thus, in order to prevail in its defense, Morgan Stanley had to prove "by the greater weight of the evidence, that it lacked knowledge of the Sunbeam fraud and did not aid and abet or conspire with Sunbeam" to defraud the plaintiff.

### **1. The Trouble With Using Keyword Searches To Identify ESI That Must Be Preserved**

As the foregoing "spoliation" cases illustrate, once litigation or a governmental investigation is reasonably anticipated, the party has a duty to identify and suspend the routine destruction of all documents or ESI that may be relevant to the anticipated litigation or investigation.<sup>100</sup> In such cases, a party must take reasonable steps to preserve information that is: (i) relevant to the action/anticipated action; (ii) reasonably calculated to lead to the discovery of admissible evidence; and (iii) reasonably likely to be requested during discovery. The first step to preserving that information is to identify where it resides.

Oftentimes, the best place to start looking for information is with the “key players” in the dispute; *i.e.*, those persons likely to have information the party will rely on to support its claims or defenses.<sup>101</sup> Counsel should interview those individuals to identify: (i) an information timeline for the dispute; (ii) what type of relevant information is likely to exist; (iii) where the person stored the information she created and/or received; (iv) what her information and data storage habits are, *e.g.*, whether she stores information on the company’s servers, a laptop, thumb drive, home computer, PDA, or one or more of the foregoing; (v) what other types of relevant information might exist concerning the dispute; (vi) who else might be a key player; and (vii) what types of information those individuals might have.<sup>102</sup> Following those interviews, counsel will also need to work with the client’s IT personnel to prepare a “data map” of where the relevant ESI is believed to be stored.

The next step is to search the company’s databases for relevant ESI and send out a litigation hold to prevent its destruction. “Keyword searches” are by far the most commonly used methodology for locating potentially relevant ESI.<sup>103</sup> The legal profession is familiar with that methodology through its use and searches of on-line legal databases.<sup>104</sup> A keyword search is a method for searching data using simple words or word combinations.<sup>105</sup> Such searches often use commands called “Boolean operators”<sup>106</sup> to expand the a search beyond the keyword root, exclude other words to limit the scope of a search, and join keywords with other terms in a way that provides added focus to a keyword search.<sup>107</sup>

Keyword searches are most often used to identify information that is responsive to discovery requests, identify privileged information, and for large-scale culling and filtering of ESI.<sup>108</sup> Keyword searches work best when the inquiry is focused on finding particular documents and the language used in those documents is relatively predictable.<sup>109</sup> In other cases, however, the vagaries of human language make keyword searches of limited value by themselves.<sup>110</sup> As one author noted, “words are living, elastic aspects of human behavior subject to constant change and only have meaning in their use.”<sup>111</sup> For instance, people in different divisions of a company or different geographic regions of a country may use different words to describe the same thing, just as people of different generations or those who are otherwise demographically distinct can essentially have their own vocabulary for discussing the same topic. Furthermore, people may make up new acronyms, words, and codes that function as a language within a language, making the task of searching for information by use of keywords especially challenging.<sup>112</sup>

The ambiguity and indeterminacy of human language means the results of a keyword search tend to be either over- or under-inclusive.<sup>113</sup> Keyword searches produce over-inclusive results when one or more of the keywords used has multiple meanings depending on the context in which the words are used. A search using the term “strike,” for example, might return documents discussing “a labor union tactic, a military action, options trading, or baseball, to name just a few.”<sup>114</sup> The problem with an over-inclusive search result is that additional time and labor will need to be expended to search the documents found for documents that are truly relevant to the issues that gave rise to the search.<sup>115</sup>

Keyword searches can return under-inclusive results when the keywords identified have overlooked synonyms or other closely-related words that are not included among the keywords used in the search. Relying on a keyword search essentially requires the parties directing the search to guess what words the authors of every document in the database to be searched would have used to express the same thought or to describe the same object or reality. For instance, “there are more than 120 words that could be used in place of the word ‘think’ (e.g., guess, surmise, anticipate).”<sup>116</sup> A keyword search might also be under-inclusive due to spelling errors, alternative ways of spelling the same word (e.g., Madeline, Madeleine, or Madelyn), or tense variations on the keyword (e.g., sing, sang, song, singing).<sup>117</sup> Finally, keyword searches that search information gathered from texts through an optical scanning process (“OCR”) are particularly likely to be under-inclusive because even the most reliable OCR processes have an error rate, meaning keywords might not be identified by the search.<sup>118</sup>

The results of a 1985 study illustrate the limits of attorneys’ abilities to effectively search and retrieve relevant documents using keywords. The study examined the effectiveness of a document production in a lawsuit involving a computerized San Francisco Bay Area Rapid Transit (“BART”) train that failed to stop at the end of the line.<sup>119</sup> The parties produced approximately 350,000 pages of documents in the case. Working with paralegals, the attorneys estimated they identified more than 75% of the relevant documents in the case. The study, however, revealed the attorneys and paralegals only found about 20% of the relevant documents. Perhaps not surprisingly, the parties used different words to search the database. For instance, the parties on the BART side of the case referred to “the unfortunate incident” whereas the parties on the victims’ side of the case used the words “accident” and “disaster” to describe the event at issue. Bolstering the premise that it is extremely difficult for attorneys to identify all keywords necessary to find all relevant documents in a case, the authors noted that for one issue the attorneys identified three keywords they believed would adequately capture the database’s relevant documents. The authors, however, discovered 26 more. Finally, the authors discovered that the terms used to discuss one of the train’s faulty parts varied depending on where in the country a document was written. The authors spent 40 hours searching for all the different terms that might be used to describe the part and gave up. As one reviewer put it, “[t]hey did not run out of alternatives, they ran out of time.”<sup>120</sup>

## **2. It May No Longer Be Reasonable To Rely On Keyword Searches Alone To Identify ESI That Must Be Preserved**

Despite the drawbacks of keyword searches, courts at one time considered keyword searches a reasonable and acceptable method for identifying ESI subject to a litigation hold.<sup>121</sup> Judge Sheindlin, author of the seminal *Zubulake* line of decisions, endorsed keyword searches developed by counsel as a reasonable means of fulfilling a party’s obligation to preserve relevant ESI.

To the extent that it may not be feasible for counsel to speak with every key player, given the size of a company or the scope of the lawsuit, counsel must be more creative. It may be possible to run a system-wide *keyword search*; counsel could then preserve a copy of each “hit.” ... Counsel does not have to review

these documents, only see that they are retained. For example, counsel could create a broad list of search terms, run a search for a limited time frame, and then segregate responsive documents ... The initial broad cut merely guarantees that relevant documents are not lost.<sup>122</sup>

Times appear to be changing. Recent federal court decisions cast doubt on whether it is reasonable for a party to rely on a keyword search, particularly one developed by counsel alone, to identify and preserve relevant information when the results of that search are not tested for under-inclusiveness through further keyword searches, sampling, or other search methods.

The court in *United States v. O'Keefe*<sup>123</sup> questioned the position endorsed by Judge Sheindlin and others that a party's use of keyword searches developed by counsel can fulfill the party's obligation to act reasonably to preserve relevant ESI. Facing a dispute over whether the government's keyword search for ESI was reasonable, the court observed that evaluating a party's search methodology "is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics."<sup>124</sup> Consequently, "for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread." The court concluded that making those determinations "is clearly beyond the ken of a layman" and, therefore, parties seeking to challenge an adversary's search methodology would need to present expert testimony to address the dispute.<sup>125</sup> As another court observed, the opinion in *O'Keefe* merely confirms that "*ipse dixit* pronouncements from lawyers unsupported by an affidavit or other showing that [a] search methodology was effective for its intended purpose are of little value to a trial judge who must decide a discovery motion aimed at either compelling a more comprehensive search or preventing one."<sup>126</sup>

In a May 2008 decision, the court in *Victor Stanley, Inc. v. Creative Pipe, Inc.*<sup>127</sup> considered the reasonableness of the defendants' use of a keyword search to conduct a privilege review of large volumes of ESI. The court was called on to determine whether the defendants' disclosure of privileged ESI following a keyword search for privileged materials was inadvertent. The issue turned on whether the defendants' decision to use a keyword search alone for their privilege review was reasonable. If it was reasonable, the disclosure could be considered inadvertent and the privilege would be preserved, but if it was unreasonable, the disclosure could not be considered inadvertent and the privileged would be deemed waived.

The parties agreed to a protocol for conducting searches for relevant ESI.<sup>128</sup> After the defendants conducted their relevancy searches, they used 70 keywords that were intended to flag all privileged materials from the ESI retrieved under the agreed search protocol.<sup>129</sup> The keywords were developed by one of the individual defendants along with defendants' counsel.<sup>130</sup> Once the keyword search was completed, the defendants produced to the plaintiff all files not flagged as a result of the keyword search.<sup>131</sup> The plaintiff discovered potentially privileged information in the materials the defendants produced and notified the defendants.<sup>132</sup> The defendants asserted that 165 of the documents produced were privileged and insisted their production of those documents was inadvertent.<sup>133</sup>

The court noted that keyword searches can be useful tools for search and retrieval of ESI but found that the results of such a search must be subjected to quality control checks to account for the limitations of that search methodology; namely, the likelihood that the search will return over- or under-inclusive results.

Common sense suggests that even a properly designed and executed keyword search may prove to be over-inclusive or under-inclusive, resulting in the identification of documents as privileged which are not, and non-privileged which, in fact, are. The only prudent way to test the reliability of the keyword search is to perform 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.<sup>134</sup>

Citing the decision in *O'Keefe*, the court also found that a party must be able to demonstrate that its keyword search was properly constructed by persons with expertise in ESI search and retrieval endeavors.<sup>135</sup> Finally, the court warned parties to be prepared to justify their chosen methodology with expert affidavits and reports demonstrating that the searches were properly implemented in the event their searches are challenged by an adversary.<sup>136</sup>

The court concluded the defendants failed to show that their keyword search was reasonable.<sup>137</sup> To begin with, the defendants did not demonstrate that the individuals who selected the keywords for the search were qualified to design an appropriate search for ESI.<sup>138</sup> The defendants further failed to show that they conducted any quality assurance testing on the results of their keyword search.<sup>139</sup> And, when the plaintiff challenged their search methodology, the defendants were wholly unprepared to explain what they had done and why it was sufficient.<sup>140</sup> As a result, the court concluded the defendants failed to show they acted reasonably to protect against the disclosure of their privileged materials. Consequently, their disclosure was deemed voluntary and the privilege waived.<sup>141</sup>

Indolent reliance on keyword searches also risks privilege waiver. In *Rhoads Indus., Inc. v. Bldg. Materials Corp. of Am.*,<sup>142</sup> the plaintiff was fortunate that the court spared it from a finding of waiver of attorney-client privilege for more than 800 privileged electronic documents inadvertently produced to the defendants. The court cited *Victor Stanley* in support of its admonition that "relying exclusively on a keyword search for the purpose of conducting a privilege review is risky, and proper quality assurance testing is a factor in whether precautions were reasonable."<sup>143</sup> The plaintiff in *Rhoads* did not test the reliability or comprehensiveness of the keyword search, and the court explicitly considered this failure as a factor weighing in favor of privilege waiver.<sup>144</sup> At very least, the plaintiff should have included the names of all of its attorneys as search terms in order to identify potentially privileged documents.<sup>145</sup>

There are two implications of the courts' decisions. First, keyword searches developed by counsel alone are probably not reasonable unless counsel has demonstrable expertise in computer sciences, statistics, and linguistics. Second, parties will need to make sure they document the searches they undertake and that the searches are defensible under the standards set by the relevant disciplines and, finally, provable in court. One author suggested living by the

accounting axiom “if it isn’t recorded, it didn’t happen” when it comes to documenting e-discovery searches,<sup>146</sup> and suggested attorneys consider various questions when deciding how to document their efforts.

For example, one part of planning for the project thus must include development of a scheme for capturing the work product generated over the life of the project. How will counsel show, if challenged, that reasonable efforts were undertaken and performed? Who will tell the story of how the work was done? What part of that story will be considered privileged, and what part available for public consideration? These kinds of questions should not arise for the first time in the middle of a project, when it may be too late to go back and document the work.<sup>147</sup>

### **3. If Keyword Searches Are Not Enough, What Search Methodologies Are Reasonable?**

The court in *Victor Stanley, Inc.* briefly touched on other search and retrieval methodologies that parties might use to search their ESI.<sup>148</sup> While a full discussion of the differing methodologies is beyond the scope of this paper, it is worth identifying some of the more prevalent methodologies and how they differ from keyword searches.

The following are specific types of searches to consider:

#### **a. Fuzzy Search Models**

So-called “fuzzy searches” look for variations of keywords by measuring the similarity of a word to the keyword. The searches score words in the set of data searched. If the word’s score meets the benchmark set by the particular test, the document in which the word is found will be retrieved. This approach can help account for common misspellings of keywords. It can also, however, produce results that are over-inclusive. For instance, a search for “Tivoli” might produce a result that includes “ravioli.”<sup>149</sup>

#### **b. Concept Search Models**

There are several types of common concept search models. These searches rely on sophisticated algorithms to evaluate whether a set of documents match a defined concept. The algorithms “essentially treat each word in a document as a number and each pattern of words as a unique series of numbers” allowing the systems to perform statistical analyses of the documents based on the definitions, frequency, and context of the words they contain.<sup>150</sup>

##### **i. Latent Semantic Analysis**

This concept search model starts with keywords and looks for other words that have a high rate of co-occurrence with other words in the data searched. This model assumes that a high rate of co-occurrence means the words are related and can be used to retrieve documents containing the co-occurring words even if it does not contain the designated keyword. For instance, in a lawsuit involving a motorcycle crash, this search method might conclude that the

words brakes, wheels, and helmet are related to the keyword “motorcycle” and retrieve documents containing those terms.<sup>151</sup>

## ii. Text Clustering

This search process uses statistical models to group together documents with similar content and displays them on a visual aid of some kind. The closer in proximity two documents are to one another on the display, the more likely they are related. Being able to review a group of related documents together makes the review more efficient. In addition, clustering can help identify documents that are not relevant to issues driving the review.<sup>152</sup>

## iii. Bayesian Classifier

This search methodology categorizes the documents in a database by taking words in a sample category and applying that rule to the other documents in the database. The process starts with a “training set” of relevant documents that serve as representative documents for the search system to look for during its search. Whether a document belongs in a particular category is a function of each word in the document and the frequency with which it appears. This search process can help to quickly identify and categorize documents as confidential, privileged, and responsive based on the training documents.<sup>153</sup>

Suffice it to say “[t]here is no one best system for all situations.”<sup>154</sup> To assist parties in evaluating and selecting the appropriate search methodology for a particular dispute, a Working Group of the Sedona Conference published the following guidelines:

- Practice Point 1:** In many settings involving electronically stored information, reliance solely on a manual search process for the purpose of finding responsive documents may be infeasible or unwarranted. In such cases, the use of automated search methods should be viewed as reasonable, valuable, and even necessary.
- Practice Point 2:** Success in using any automated search method or technology will be enhanced by a well-thought out process with substantial human input on the front end.
- Practice Point 3:** The choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed.
- Practice Point 4:** Parties should perform due diligence in choosing a particular information retrieval product or service from a vendor.

- Practice Point 5:** The use of search and information retrieval tools does not guarantee that all responsive documents will be identified in large data collections, due to characteristics of human language. Moreover, differing search methods may produce differing results, subject to a measure of statistical variation inherent in the science of information retrieval.
- Practice Point 6:** Parties should make a good faith attempt to collaborate on the use of particular search and information retrieval methods, tools and protocols (including as to keywords, concepts, and other types of search parameters).
- Practice Point 7:** Parties should expect that their choice of search methodology will need to be explained, either formally or informally, in subsequent legal contexts (including in depositions, evidentiary proceedings, and trials).
- Practice Point 8:** Parties and the courts should be alert to new and evolving search and information retrieval methods.<sup>155</sup>

#### **D. The Need For And Selection Of Third-Party E-Discovery Vendors**

There are several general stages to the process of producing documents and ESI during discovery: (1) identification; (2) preservation; (3) collection of information; (4) processing; (5) review for privilege; (6) analysis for responsiveness; and (7) production of responsive information.

The costs of e-discovery along with improvements in software that can search, retrieve, and prevent destruction of ESI across an enterprise's disparate systems have caused some companies to look at bringing at least some of their e-discovery processes in-house.<sup>156</sup> Cisco Systems, Inc. elected to bring its e-discovery process in-house before most. It created a unique process that made the collection, processing, and outside-counsel review far more efficient and less costly. Cisco adopted the process after facing a staggering \$23,500,000 bill for e-discovery – and that was the bill for just *one lawsuit* that followed the 2000-2001 stock market decline.<sup>157</sup> Cisco described the company's process as "build[ing] a mountain of information and tear[ing] it down."<sup>158</sup> The process starts with Cisco gathering information "from literally thousands of sources" across the company and then copying all of the retrieved data to a litigation repository.<sup>159</sup> Once in the repository, Cisco prepares the files for legal review by:

- (1) Converting e-mail messages "into separate files and paired with their attachments";
- (2) Removing irrelevant application files, help files, "read-me" files, and "log files," (which can reduce the volume of data to review by up to 70 percent);

- (3) Scanning the files for duplicates, which are tagged and removed from the repository (which also reduces data volume); and
- (4) Assigning "a unique document number" to each file "for ease of review and retrieval."<sup>160</sup>

Once that process is complete, Cisco makes the data available on-line for review by outside counsel.<sup>161</sup> Counsel then electronically marks the data as relevant or irrelevant and separates proprietary and privileged information for special handling.<sup>162</sup> The remaining data is broken down into separate files, converted to a TIFF format, burned to CDs or DVDs and given to opposing counsel for review.<sup>163</sup> Cisco reports that its effort has been hugely successful, reducing expenditures for outside counsel legal review by 30 percent and reduced its overall discovery costs by 97 percent.<sup>164</sup>

While the results are enviable, Cisco's investment in this type of process is probably a luxury that most companies cannot afford or even justify based on current e-discovery needs. For that reason, many companies opt for bringing some e-discovery functions in-house while continuing to rely on third-party vendors for a significant portion of their e-discovery needs.<sup>165</sup>

Using a third-party vendor serves functional and strategic purposes. On the functional side, a vendor can help where the scope of a project is too complex or burdensome to handle in-house, or where a company needs specialized equipment or knowledge in order to access, process, and produce the relevant ESI. Strategically, a company might use vendors to ensure the defensibility of its e-discovery process. Retaining a vendor whose search methodologies and processes have been challenged and prevailed in other courts can help establish a company's "due diligence" for preserving and producing ESI. "However, hiring a vendor does not absolve an attorney from responsibility. The attorney retains the obligation to oversee and direct the management of the ESI. As a bottom line, the vendor operates at the direction of the attorney; thus there are ethical considerations for having ESI management outsourced."<sup>166</sup> Accordingly, "[h]aving an indemnification provision in the agreement with the vendor may ultimately protect the attorney financially, but it does not supplant their responsibility."<sup>167</sup>

The industry of e-discovery vendors continues to develop. Selecting third-party vendors can be a complex process. Thus, in 2007, the Sedona Conference published guidelines for selecting vendors in "Best Practices for the Selection of Electronic Discovery Vendors: Navigating the Vendor Proposal Process."<sup>168</sup> The Sedona Conference concluded that companies will benefit from using a systematic process for comparing vendors. As set forth in those guidelines, a company should identify the scope and type of services it needs.<sup>169</sup> It should then identify the vendors who provide the needed services, perform some general research, and formally request information from the most promising options.<sup>170</sup> After receiving the requested information, the company can compare the vendors, further narrow its options, and solicit proposals from the vendors it determines will best meet its needs.<sup>171</sup>

## 1. Define The Services Required

It is imperative to determine the scope of the company's needs as precisely as possible. A company cannot possibly determine what vendor should be handling its e-discovery project unless the company defines exactly what that project will entail.<sup>172</sup> For example, the company might need a vendor to recover data from outdated legacy systems, harvest data and e-mail from sources throughout the company, and process and prepare them for production.<sup>173</sup> In contrast, the company might simply need an outside consultant to review its IT infrastructure, determine where relevant data may be stored, and provide a plan for preservation and/or production of that material.<sup>174</sup> Counsel should "estimate the size and complexity of the project," including "the volume of data," and the number of "servers and custodians."<sup>175</sup>

### a. Categories Of Vendor Services

The Sedona Conference sets forth five general categories of services e-discovery vendors provide: "1) Consulting/Professional Services; 2) Data Collection/Processing; 3) Data Recovery/Forensics; 4) Hosting/Review/Production/Delivery; and 5) Other Litigation Support-Related Services."<sup>176</sup> While vendors will typically provide a combination of such services, they often specialize in one area.<sup>177</sup> The company should identify vendors who specialize in its area of need, and select several potential candidates.

The following is a summary of the types of services available in each of the five general categories. Consultants can provide an overall analysis of the company's e-discovery issues, analyze its current document retention policies and the status of its information infrastructure, and make recommendations about how the company should handle electronic data both in general, and in the context of pending litigation.<sup>178</sup> Data Collection vendors specialize in gathering electronic data from various sources, filtering it, removing metadata, redacting any necessary portions, and converting it into the agreed-upon form for production.<sup>179</sup> Previously, this generally meant converting data into .pdf files. More recently, however, some companies opt to produce electronic data in its native format, which can be more efficient.<sup>180</sup> Data Recovery entails the restoration of data stored in an inaccessible format, such as a "legacy" system no longer in use, or on damaged or corrupted media. Often, this data cannot be accessed without the use of specialized equipment.<sup>181</sup> Data Hosting services store data and make it available on an ongoing basis (for example, on a website).<sup>182</sup> A data hosting company can collect and store electronic data in an accessible fashion prior to any litigation and to make certain documents available to all parties via a secure internet server during litigation.<sup>183</sup> Other Litigation Support services include, for example, document scanning or the conversion of paper documents into searchable electronic text data.<sup>184</sup>

### b. Requesting Information

Once a company has determined the types of services it will require and has identified several potential vendors, it should gather information about them in order to refine its

options.<sup>185</sup> “With the scope of the project in mind, counsel should identify three or more service providers with varying qualifications to submit proposals for the vendor portion of the work.”<sup>186</sup>

## **2. Examining Potential Vendors**

### **a. Qualifications And Reputation**

The general business reputation of any e-discovery vendor is always crucial.<sup>187</sup> When selecting a vendor, a company should consider the vendor’s reputation for integrity and sound business practices, how long it has been in business, how many employees it has, whether the project will be handled entirely in-house, and whether the vendor will use sub-contractors for any portion of the project.<sup>188</sup> In responding to a request for information, vendors should be able to provide references, including clients for whom they have worked in the past.<sup>189</sup> The company should also inquire as to the vendor’s history and financial background and its experience in handling similar projects, including any expert testimony defending its e-discovery processes that the vendor has provided on behalf of a client in the past.<sup>190</sup> The vendor should be able to clearly explain its process for handling e-discovery projects, provide a general timeline for completion of the project, and provide information regarding the qualifications and experience of the individual employees who would be handling the project.<sup>191</sup> Further, “[t]he vendor should assign at least one permanently dedicated project manager to the case,” and establish “[q]uality control systems.”<sup>192</sup>

### **b. Security**

Even if the outside vendor is qualified and reputable, it must demonstrate it is capable of keeping its clients’ information secure.<sup>193</sup> The Sedona Conference recommends visiting the vendor’s offices to “kick the tires” and review the vendor’s security measures.<sup>194</sup> An e-discovery vendor should have a system for restricting physical access to its network hardware to the necessary employees and for preventing the removal of data from the premises, as well as a system to prevent access by hackers or corruption by computer viruses.<sup>195</sup> In addition, the vendor should have a robust backup system in place, preferably including off-site backup.<sup>196</sup> The vendor should also demonstrate that its employees are trusted and reliable.<sup>197</sup> Specifically, a company should inquire whether the vendor performs background checks on its employees and requires employees who leave the vendor to sign confidentiality agreements.<sup>198</sup> In addition, a company should determine how the vendor will handle the data when the project is complete.<sup>199</sup>

### **c. Conflicts**

As with any business relationship, conflicts are possible. A company must determine whether a vendor is working with or previously has worked with its business competitors or adversary in the litigation.<sup>200</sup> Also, in completing an e-discovery project, a vendor will probably receive confidential information regarding a company’s business practices, trade secrets, or litigation strategies. These additional concerns and potential conflicts must be considered as well prior to hiring an e-discovery vendor.<sup>201</sup> In some cases, the nature of the conflict will be acceptable, such as where the vendor performed copying services for a competitor. In other

cases, the conflict will necessarily eliminate a vendor from consideration, regardless of its other qualifications.<sup>202</sup>

The conflicts investigation is an ongoing process. A vendor may wish to work with a company's competitors in the future, creating additional potential confidentiality issues not present at the outset of the relationship.<sup>203</sup> Moreover, because the e-discovery industry remains in its relative infancy, a vendor may acquire or merge with another vendor that may have a company's competitors or adversaries in litigation as clients.<sup>204</sup> Accordingly, the Sedona Conference recommends including a provision regarding conflicts in any services agreement, which sets forth the parties' rights and duties regarding potential conflicts and provides that the vendor will not work for the company's business competitors or legal adversaries.<sup>205</sup>

#### **d. Pricing Models**

Accurately determining the total cost of an e-discovery project can be daunting, and prices vary with each e-discovery vendor.<sup>206</sup> A company must fully understand the vendor's pricing system, including all potential additional costs and areas of potential cost overruns, before it selects a vendor.<sup>207</sup> "While the vast majority of all electronic data was traditionally converted (to TIFF, PDF or HTML, for example) for review and production (either on paper or in load files), it is becoming more prevalent for vendors to offer processes allowing the review to take place in 'native' format."<sup>208</sup> Because the data was being converted into page format, those e-discovery vendors used a "per-page" system for pricing.<sup>209</sup> Because data in a native digital format may not be in the form of "pages," a per-page pricing model is insufficient. Accordingly, some vendors use a "per-gigabyte" pricing model.<sup>210</sup>

A company seeking to hire an e-discovery vendor should confirm that the vendor uses the most efficient methods possible. For example, rather than reviewing entire hard disks, the vendor should be able to eliminate system files and redundant data.<sup>211</sup> As explained above, parties often discuss and agree on the scope of electronic data that must be reviewed.<sup>212</sup> Armed with such an agreement, a vendor can further limit the scope of its inquiry (and the associated costs) by restructuring its search to a specific time period, or even to particular directories and disks most likely to contain relevant data.<sup>213</sup>

### **3. Soliciting Proposals And Making The Selection**

After the company reviews the information received from its potential vendors and refines its options, it can solicit formal proposals for its e-discovery project from the vendors.<sup>214</sup> The company should provide as thorough a description of the project as possible, describing the general type of information that is being sought, its sources of electronic data and how that data is stored, and the timelines (and court deadlines) for production of that data.<sup>215</sup> The company should articulate its expectations for the project in terms of the roles it and the vendor will serve, the expected benchmarks for completion of the project, and the frequency and form of expected status reports.<sup>216</sup> The proposal request should also indicate the e-discovery services that will be required and the processes that the company expects to be used.<sup>217</sup> Of course, the request should also allow the vendor to identify any alternative methods or processes that it believes may be

