# Searching for an Answer: Defensible E-Discovery Search Techniques in the Absence of Judicial Voice

*Harrison M. Brown*[*]

"No longer can the time-honored cry of 'fishing expedition' serve to preclude a party from inquiring into the facts underlying his opponent's case. Mutual knowledge of all the relevant facts gathered by both parties is essential to proper litigation."[1]

The past two decades have seen a widespread shift from original physical information storage technologies, to new, digital information technologies, resulting in an exponential rise in the amount of information that is created, processed, and stored.[2] This "inflationary dynamic" has caused written information to increase to never-before-seen levels, resulting in a new landscape which makes it prohibitively expensive, if not impossible, for litigation to carry on as it has up until now.[3]

"Today, most litigation includes electronically stored information (ESI)[4] as a critical aspect of the discovery and production phase."[5] Because

---

* J.D. Candidate May 2013, Chapman University School of Law; B.A. Political Science and History 2010, University of California, Los Angeles. I wish to thank my parents, Richard and Ann, and my sister, Rebecca, for their encouragement in this project and throughout law school. Thank you to the *Chapman Law Review* members for their assistance on this article and their commitment to the Journal. Special thanks to Professor John Hall for his support in this endeavor.

1 Hickman v. Taylor, 329 U.S. 495, 507 (1947).

2 *See* GEORGE L. PAUL & BRUCE NEARON, THE DISCOVERY REVOLUTION: E-DISCOVERY AMENDMENTS TO THE FEDERAL RULES OF CIVIL PROCEDURE 1 (2006) ("Society stores information in a profoundly different way than it did in 1990.").

3 *See* George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J.L. & TECH. 10, 1–2 (2007), http://law.richmond.edu/jolt/v13i3/article10.pdf (deriving the term "inflationary dynamic" from ALAN H. GUTH, THE INFLATIONARY UNIVERSE: THE QUEST FOR A NEW THEORY OF COSMIC ORIGINS (1997)).

4 "ESI includes e-mails, webpages, word processing files, and databases stored in the memory of computers, magnetic disks (such as DVDs and CDs), and flash memory (such as 'thumb' or 'flash' drives)." BARBARA J. ROTHSTEIN ET AL., FED. JUDICIAL CTR., MANAGING DISCOVERY OF ELECTRONIC INFORMATION: A POCKET GUIDE FOR JUDGES 2 (2007), *available at* http://www.fjc.gov/public/pdf.nsf/lookup/eldscpkt.pdf/$file/eldscpkt.pdf; *see also* HANDBOOK OF DIGITAL FORENSICS AND INVESTIGATION 63–64 (Eoghan Casey ed.) (2009) (distinguishing between "e-discovery," defined as the "exchange of data between parties in civil or criminal litigation," and "ESI," which is the electronic data that itself is the subject of litigation).

5 *Is 'Manual' Collection of ESI Defensible?* INT'L ASS'N FOR INFO. MGMT. PROF'LS (Apr. 10, 2010), http://www.arma.org/news/enewsletters/index.cfm?ID=4270.

ESI is produced in such large quantities and the increase in ESI easily adds to the cost of review, manual or linear review has significantly decreased in e-discovery cases.[6] In its place, attorneys have frequently used legacy search techniques, such as keyword searches,[7] to filter data for producing responsive documents in discovery.[8] These search methods, however, are not without their own problems and are increasingly coming under attack.[9]

Instead, advanced automated search methods such as concept searching[10] and predictive coding[11] have emerged as efficient ways to comb ESI for responsive documents and are "more likely to produce the most comprehensive results."[12] Although progress has been made in recent

---

6 *See, e.g.*, Herbert L. Roitblat et al., *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. AM. SOC'Y. FOR INFO. SCI. AND TECH. 70, 70 (2010) (advising that exhaustive manual review, conducted linearly, requires one or more persons to examine each document in a collection and to code it as responsive or non-responsive); *see also* Sedona Conference Working Grp. on Best Practices for Document Retention and Prod., *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189, 194–195 (2007) [hereinafter *Best Practices Commentary*] ("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. . . . A consensus is forming in the legal community that [manual] review of documents in discovery is expensive, time-consuming, and error-prone."); *see also infra* Part I. Electronic discovery ("e-discovery") refers to discovery of documents produced in electronic formats rather than hardcopy for litigation. *Definition of Electronic Discovery*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/glossaries/glossary/e/electronic-discovery (last visited Oct. 18, 2012).

7

> A *keyword search* is a basic search technique that involves searching for one or more words within a collection of documents. Typically, a keyword search involves a user typing their search request, or query, into a search engine . . . which then returns only those documents that contain the search terms entered.

*Search Methodologies*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/guides/edrm-search-guide/search-methodologies (last visited Nov. 27, 2011); *see also infra* Part II.D.

8 *See* In re Lorazepam & Clorazepate, 300 F. Supp. 2d 43, 47 (D.D.C. 2004) (requiring the use of keyword search terms as a reasonable means of narrowing production in e-discovery).

9 *See infra* Part II.D.

10 "Concept search allows a legal professional to specify a concept and documents that describe that concept to be returned as the search results . . . . Concept search solutions rely on sophisticated algorithms to evaluate whether a certain set of documents match a concept." *Search Methodologies*, *supra* note 7; *see also infra* Part III.A.

11 Predictive coding is

> a combination of technologies and processes in which decisions pertaining to the responsiveness of records gathered or preserved for potential production purposes … are made by having reviewers examine a subset of the collection and having the decisions on those documents propagated to the rest of the collection without reviewers examining each record.

*E-Discovery Institute Survey on Predictive Coding*, ELEC. DISCOVERY INST., 2 (Oct. 1, 2010), http://www.ediscoveryinstitute.org/pubs/PredictiveCodingSurvey.pdf; *see also infra* Part III.B.

12 *See* Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139, 148 (D.D.C. 2007) (suggesting the parties consider "concept searching, as opposed to keyword searching"); *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, 3 (2011), http://law.richmond.edu/jolt/v17i3/article11.pdf ("[A] technology-assisted process, in which humans examine only a small fraction of the document collection, can yield higher recall and/or precision than an exhaustive manual review process . . . .").

years, many attorneys remain reluctant to move away from less reliable manual review and legacy search methods and embrace advanced search techniques; this is in part due to a lack of consensus on which particular technology should be used.[13] While the bench is at times supportive of advanced search techniques,[14] it has yet to expressly endorse one type.[15]

Rather than wait for judicial approval of a particular kind of technology, which may not come, counsel should cooperate throughout the entire process of electronic discovery.[16] Cooperating with opposing counsel in developing search protocols will help avoid disputes that may later arise about the appropriateness and sufficiency of search efforts taken by each party, which in turn will reduce discovery deficiencies.[17] Developing and documenting a defensible search methodology prepares a party to defend the reasonableness of search protocols should a dispute arise and assures quality control in e-discovery.[18]

Part I of this Note describes the modern information inflationary epoch and how traditional manual document review and production cannot keep pace with the demands inherent in this sea of change. Part II surveys institutional attempts to streamline e-discovery and investigates the efficacy of commonly used legacy search methodologies. Part III introduces two of the most promising alternative search techniques in practice today. Part IV examines recent case law and other authorities on whether e-discovery experts are needed to support a party's search protocols. Lastly, Part V discusses steps parties can take to create defensible search protocols in light of the bench's silence on its preferred search methodologies.

## I. MANUAL REVIEW—FROM GOLD STANDARD TO OBSOLETE

Although the way people communicated through written media remained unchanged for many years, the world has recently seen evolutionary changes in the way people write and communicate.[19] This shift is primarily a result of the advent of the personal computer as well as

---

13 *See* WILLIAM WEBBER, RE-EXAMINING THE EFFECTIVENESS OF MANUAL REVIEW 8 (2011), *available at* http://www.umiacs.umd.edu/~wew/papers/w11sire.pdf (noting that it still remains uncertain which method can most thoroughly and reliably meet supervising attorneys' document review goals).

14 *See, e.g.* William A. Gross Const. Assocs. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 134–36 (S.D.N.Y. 2009)("strongly" endorsing the Sedona Conference methods of ESI retrieval and admonishing attorneys for using "seat of the pants" methods instead).

15 *See* Victor Stanley, Inc. v. Creative Pipe, Inc. 250 F.R.D. 251, 259 n.9 (D. Md. 2008) (discussing how alternative electronic search methods "can enhance the accuracy and reliability of [a] search," but not going so far as to offer a preference for a particular type of search method).

16 *See infra* Part VI.

17 *See infra* Part VI; *see also Gross*, 256 F.R.D. at 136.

18 *See infra* Part VI.

19 Information technology remained "simple" and in "equilibrium for over 5200 years;" however, advances in technology have quickly lead to "an evolutionary burst in writing technology." Paul & Baron, *supra* note 3, at 4–5.

the growth of interconnected global networks.[20] Consequently, the total amount of written information has multiplied to previously unimaginable levels.[21] This growth in volume has had a profound impact on litigation as "it places at severe risk the justice system's ability to achieve the 'just, speedy and inexpensive' resolution of disputes, as contemplated by Rule 1 of the Federal Rules of Civil Procedure."[22] As such, manual review, once considered the "gold standard" of document review,[23] is now infeasible and obsolete in an increasing number of cases.[24]

A. Information Inflation

Information technology, simple and static for more than fifty centuries, has drastically changed in recent years as an evolution in writing resulted in information inflation. This is primarily attributable to the emergence of a "'digital realm' . . . created by an accretion of technological advances, each built on preceding advances."[25] These advances "include digitization; real time computing; the microprocessor; the personal computer, email; local and wide-area networks . . . the evolution of software . . . [and] the World Wide Web . . . ."[26]

The past two decades have seen an exponential rise in the amount of information that is created, processed, and stored. "Computers have enabled the [large-scale] creation of [] information . . . and unleashed an unprecedented deluge of data," [27] the results of which are staggering.[28] In

---

20 PAUL & NEARON, *supra* note 2, at 2–3.

21 *See* Paul & Baron, *supra* note 3, at 1–2 n. 2 (noting that "[o]rganizations now have thousands if not tens of thousands of times as much information within their boundaries as they did 20 years ago" (quoting PAUL & NEARON, *supra* note 2, at 4)).

22 *Best Practices Commentary*, *supra* note 6, at 197; *see also* Fed. R. Civ. P. 1 (2012).

23 *Best Practices Commentary*, *supra* note 6, at 199 (

> [T]here appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible—perhaps even perfect—and constitutes the gold standard by which all searches should be measured . . . . [However], the relative efficacy of that approach versus utilizing newly developed automated methods of review remains very much open to debate.)

24 *See, e.g.*, Robert W. Trenchard & Steven Berrent, *Can Technology 'De-Commoditize' Document Review?*, LAW TECH. NEWS (Apr. 28, 2011), http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202491954188&slreturn=1 ("[I]n the modern age, when computers create and retain far more information than was ever before thought possible, the old model of manual document review is becoming increasingly unworkable.").

25 *Best Practices Commentary*, *supra* note 6, at 197.

26 Paul & Baron, *supra* note 3, at 5–6.

27 Bennett B. Borden, *The Demise of Linear Review*, ST037 ALI-ABA 277, 279 (2011).

28 "By 2012, 20 typical broadband households will generate more traffic than flowed across the entire internet in 2008." Dave Evans & Rick Hutley, *The Explosion of Data: How to Make Better Business Decisions by Turning "Infolution" Into Knowledge*, CISCO INTERNET BUS. SOLUTIONS GRP., 1 (2010), http://cco.cisco.com/web/about/ac79/docs/pov/Data_Explosion_IBSG.pdf. Evans and Hutley also noted that an amount of digital data equivalent to the entire Library of Congress is created every five minutes. *Id.*; *see also* John F. Gantz et al., *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011*, INT'L DATA CORP. (Mar. 2008), http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf.

2006 alone, the world "created, captured and replicated enough digital information to fill all of the books ever created in the world, 3 million times."[29] Society simply stores information in a profoundly different way than it did previously.[30] Because of advances in technology and the integration of society into cyber-networks, the world has been forced to adapt to an ever-changing digital frontier.

In the legal world, the various types of discoverable materials in digital form are proliferating. ESI covers data similar to previous hard-copy documents, but also includes more types that were never found in the pre-electronic world, such as e-mail messages.[31] An estimated 247 billion e-mail messages were sent in 2009, a number expected to more than double by 2013.[32] As of 2010, the average corporate worker sends and receives upwards of 110 e-mail messages per day.[33] Other types of information now discoverable as ESI include "instant messaging, word processing with hyperlinks, integrated voice mail, . . . structured databases of all kinds, Web pages, blogs, and e-data in all conceivable forms."[34] With the types and volume of ESI continuing to expand to enormous levels, the use of manual review as a viable tool in litigation is seemingly in doubt.

### B. Manual Review is Ill-Suited for Today's Legal World

The traditional "discovery review process is poorly adapted to much of today's litigation."[35] Manual review is being forced out of the litigation process as a result of time constraints and skyrocketing costs associated with the information inflation.[36] With the amount of ESI in lawsuits expanding greatly, "[t]he cost of manual review . . . is prohibitive, often exceeding the damages at stake."[37] Moreover, large data sets often make it impossible to complete manual review in a timely manner.[38] Lastly, the efficacy of manual review has been greatly called into question.[39]

---

29 *The Sedona Conference Commentary on ESI Evidence and Admissibility*, SEDONA CONFERENCE WORKING GRP. ON ELEC. DOCUMENT RETENTION AND PROD. (Mar. 2008), *available at* http://www.thesedonaconference.org/dltForm?did=ESI_Commentary_0308.pdf.

30 *See* PAUL & NEARON, *supra* note 2, at 2–3.

31 *See* JAMES N. DERTOUZOS ET AL., RAND INST. FOR CIVIL JUSTICE, THE LEGAL AND ECONOMIC IMPLICATIONS OF ELECTRONIC DISCOVERY: OPTIONS FOR FUTURE RESEARCH 1 (2008), *available at* http://www.rand.org/pubs/occasional_papers/2008/RAND_OP183.sum.pdf.

32 Masha Khmartseva, *Email Statistics Report*, RADICATI GRP., 2009–2013*,* 3 (Sara Radicati ed., May 2009), http://www.radicati.com/wp/wp-content/uploads/2009/05/email-stats-report-exec-summary.pdf.

33 *Email Statistics Report*, RADICATI GRP., 3 (Sara Radicati ed., Apr. 2010), http://www.radicati.com/wp/wp-content/uploads/2010/04/Email-Statistics-Report-2010-2014-Executive-Summary2.pdf.

34 *See* Paul & Baron, *supra* note 3, at 14.

35 *See Best Practices Commentary*, *supra* note 6, at 198.

36 *Id.*

37 *Id.*

38 *Id.*

39 *Id.*

### C. Manual Review Cannot Keep Up With the Demands of Modern Litigation

The huge volume of available ESI poses unique challenges—both in terms of cost and time to complete the review—which traditional document review simply cannot meet. Prior to the recent information inflation, complying with discovery requests evoked a familiar image of young attorneys wading through "mountains of boxes filled with dusty, poorly organized documents."[40] Confronted with such a task, the only practical action that could be taken was to read each document linearly, or in a serial fashion.[41]

While the presence of hundreds of boxes of documents may have been concerning to young associates just a few years ago, today that same amount of data might be found on a single computer hard drive.[42] Additionally, as the ability to create and store copious amounts of data rapidly increases, the cost to store that information falls.[43] Consequently, "more individuals and companies are generating, receiving and storing more data, which means more information must be gathered, considered, reviewed and produced in litigation."[44] Whereas a small business may have once had a single file cabinet full of paper records, a typical small business today stores the digital equivalent of as many as 2,000 file cabinets.[45]

Accordingly, manual review is becoming neither workable nor economically feasible. As the court remarked in *Pension Committee v. Banc of America*, we live in "an era where vast amounts of electronic information is available for review," and therefore "discovery in certain cases has become increasingly complex and expensive."[46] E-discovery accounts for as much as 25% of the total cost of litigation, and the biggest single cost in the process is attorney review time of voluminous data.[47] "[T]o the extent that a particular document is likely to be the object of a discovery request, it potentially can also represent a very real liability. The cost of collection, review and production often exceeds $2 per document—

---

40 Borden, *supra* note 27, at 279.

41 *Id.*

42 "[O]ne gigabyte of electronic information can generate approximately 70,000–80,000 of text pages, or 35 to 40 banker's boxes of documents (at 2,000 pages per box). Thus, a 100-gigabyte storage device . . . could hold as much as the equivalent of 3,500 to 4,000 banker's boxes of documents." *Best Practices Commentary*, *supra* note 6, at 192 n.2.

43 "Over the last 30 years, space per unit cost has doubled roughly every 14 months." Matthew Komorowski, *A History of Storage Cost* (July 24, 2009), http:// http://www.mkomo.com/cost-per-gigabyte (emphasis omitted). Whereas a five megabyte (MB) hard drive cost as much as $3,500.00 in 1981 (the equivalent of $700,000.00 per gigabyte (GB)), a modern hard drive retails for less than $0.10 per GB. *Id.*; *see also* John Gantz & David Reinsel, *Extracting Value From Chaos*, INT'L DATA CORP., 4 (June 2011), http://idcdocserv.com/1142 (showing a projected decrease in cost per GB from 2005 to 2015).

44 *Best Practices Commentary*, *supra* note 6, at 192.

45 *See* PAUL & NEARON, *supra* note 2, at 4–5.

46 Pension Comm. of Univ. of Montreal Pension Plan v. Banc of Am. Sec., LLC, 685 F. Supp. 2d 456, 461 (S.D.N.Y. 2010).

47 *See* Roitblat et al., *supra* note 6, at 70.

and corporations produce and store many billions of documents annually."[48] As such, it is not unusual for the cost of reviewing information to exceed the damages at stake,[49] forcing companies to settle cases out of necessity, rather than based on the merits.[50]

Moreover, large amounts of ESI make it impossible to meet the time constraints imposed in litigation. For example, it would take approximately fifty-four years to complete the review of a dispute with one billion e-mails, with one hundred reviewers working ten hours per day, seven days a week.[51] Limiting review to just one percent of the total universe of documents would still take twenty-eight weeks to complete.[52]

This scenario is increasingly becoming a reality, as seen recently in *In Re Fannie Mae Securities Litigation*, where the D.C. Court of Appeals affirmed the district court's order holding the Office of Federal Housing Enterprise Oversight (OFHEO)—the federal agency that regulates Fannie Mae and Freddie Mac—in contempt for failing to comply with a discovery deadline to which it agreed.[53] In 2006, individual defendants who were former Fannie Mae executives subpoenaed thirty categories of documents from OFHEO, a nonparty to the litigation.[54] In 2007, after the OFHEO claimed that it had produced all the documents requested, the defendants later conducted a Rule 30(b)(6) deposition[55] of OFHEO and learned that OFHEO had failed to search all of its off-site records.[56] Later, after OFHEO failed to produce additional documents, the individual defendants moved to hold OFHEO in contempt.[57] After the contempt hearing began, the parties stipulated that OFHEO would continue to conduct searches and provide all responsive documents by January 2008.[58]

Requiring them to review approximately 660,000 documents, "OFHEO undertook extensive efforts to comply with the stipulated order, hiring [fifty] contract attorneys solely for that purpose. The total amount OFHEO spent on the individual defendants' discovery requests eventually

---

48 Borden, *supra* note 27, at 279; *see also Best Practices Commentary*, *supra* note 6, at 198 n.13 (noting that for an associate reviewing an average of fifty documents at ten pages in length each hour, it would take the associate 160 hours to review one gigabyte of data at a billable rate of $200 per hour, for a total cost of $32,000).

49 *Best Practice Commentary, supra* note 6, at 198.

50 *See* Steven Hunter, *E-Discovery: Cutting Costs With Predictive Coding*, INSIDE COUNSEL (Sept. 7, 2011), http://www.insidecounsel.com/2011/09/07/e-discovery-cutting-costs-with-predictive-coding.

51 Jason R. Baron & Michael D. Berman, *Designing a "Reasonable" E-Discovery Search: A Guide for the Perplexed*, *in* MANAGING E-DISCOVERY AND ESI: FROM PRE-LITIGATION THROUGH TRIAL 479, 481 (Berman et al. eds., 2011).

52 *Id.*

53 In re Fannie Mae Sec. Litig., 552 F.3d 814 (D.C. Cir. 2009).

54 *Id.* at 816.

55 Fed. R. Civ. P. 30(b)(6) (2012) (providing for depositions of adverse organizations through designated representatives).

56 *In re Fannie Mae*, 552 F.3d at 817.

57 *Id.*

58 *Id.*

reached over $6 million, more than 9 percent of the agency's entire annual budget."[59] Despite this, after moving for and receiving two extensions, OFHEO failed to meet the deadline.[60] The district court granted the individual defendants' renewed motions for contempt, finding that "OFHEO's efforts at compliance were 'not only legally insufficient, but too little too late.'"[61] The district court imposed sanctions on OFHEO, and the Court of Appeals upheld the sanctions.[62]

*Fannie Mae* highlights the problem with manual review: parties using this method will have to commit time and resources that are simply not available. The volume and associated complexity in having to search through large amounts of ESI will only worsen as time goes on, and manual review is ill-equipped to confront the problem. As such, "automated search methods should be viewed as reasonable, valuable, and even necessary."[63]

## II. THE MYTH OF MANUAL REVIEW AS THE GOLD STANDARD IN DISCOVERY

Prior to the information inflation, manual review was long considered the "gold standard" in discovery.[64] However, as discussed above, manual review is increasingly becoming more challenging by the sheer amount of data typically generated and stored by almost every organization that uses computer technology.[65] Even assuming, arguendo, that practitioners had the resources and time to undertake manual review of voluminous sets of ESI, studies demonstrate that manual review of large data sets is imprecise and fails to live up to its billing.[66]

A widely-cited study on the efficacy of manual review, conducted by David Blair and M.E. Maron in 1985, shows the problems inherent in the use of human language among the various persons who can be involved in a dispute, and how difficult it can be to take this into account in a search for informational records.[67] The Blair and Maron study involved a manual review of about 40,000 documents spanning 350,000 pages of text captured in an IBM database to be used in a large corporate lawsuit.[68] Attorneys collaborated with paralegal search specialists to find all of the relevant

---

59 *Id.*

60 *Id.* at 817–18.

61 *Id.* at 818.

62 *Id.* at 823–24.

63 *Best Practices Commentary*, *supra* note 6, at 194.

64 *See supra* text accompanying note 23.

65 *See supra* Part I.B.

66 *See, e.g.*, Grossman & Cormack, *supra* note 12, at 3; *see also 2010 Legal Track Results*, UNIV. OF MD. INST. FOR ADVANCED COMPUTER STUDIES, http://trec-legal.umiacs.umd.edu/#2010 (last visited Nov. 23, 2011); *see also* Roitblat et al., *supra* note 6, at 72.

67 *See generally* David C. Blair & M. E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, 28 COMMC'NS OF THE ACM 289 (1985).

68 *Id.* at 290–91.

documents.[69] The attorneys estimated that they had found 75.5% of the relevant documents, however Blair and Maron's more detailed analysis found that the actual recall value was 20.26%, meaning that the attorneys believed that they were retrieving a much higher percentage of relevant documents than they actually were.[70]

Blair and Maron found that the different parties in the case used different words in their search for relevant documents, depending on their point of view.[71] For example, the attorneys representing "[t]hose who were personally involved in the event, and perhaps culpable, tended to refer to it euphemistically as, inter alia, an 'unfortunate situation,' or a 'difficulty.'"[72] However, "[t]hose who discussed the event in a critical or accusatory way referred to it quite directly—as an 'accident.'"[73]

Blair and Maron also found that the efficacy of manual review is directly tied to the amount of documents to be evaluated.[74] Notably, they found that "the value of Recall decreases as the size of the database increases, or, from a different point of view, the amount of search effort required to obtain the same Recall level increases as the database increases, often at a faster rate than the increase in database size."[75] Thus, manual review is plagued not only by time and expense constraints, but becomes a less effective tool as document universes grow, making it ill-suited for modern litigation.

The information inflation we are experiencing as a result of the incorporation of technology and new tools in society presents new challenges to litigation. While technology may be the source of the problem in e-discovery, it also appears to be the best possible solution.[76] As lawyers began to realize that manual review could not keep pace with the demands of e-discovery, the legal community began to collaborate and establish working models to assist in the discovery process.[77] Attorneys also utilize tools like optical character recognition (OCR)[78] technology to digitize paper documents in order to make use of search methodologies

---

69 *Id.* at 291.

70 *Id.* at 293.

71 *Id.* at 295.

72 *Id.*

73 *Id.*

74 *Id.* at 296.

75 *Id.* "Recall measures how well a system retrieves *all* the relevant documents" in a given set, and is represented as a proportion of relevant documents retried by a system. *Id.* at 290.

76 *Best Practices Commentary*, *supra* note 6, at 192.

77 *See id*.

78 "Optical character recognition is the conversion of a scanned document into searchable text and the rendering of its text" into a form that a computer can manipulate. *Definition of Optical Character Recognition*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/glossary/o/ocr (last visited Dec. 2, 2011). Printed material is scanned and converted into electronic files that "can then be searched for specific words or phrases." *Id.*

such as keyword and Boolean searches that increase efficiency and reduce costs.[79]

A. Attempting to Bring Order to a Disorderly Problem

E-discovery experts and consultants George Socha and Tom Gelbmann co-founded the Electronic Discovery Reference Model ("EDRM") in "May 2005 to address the lack of standards and guidelines in the e-discovery market."[80] Since then, "over 900 e-discovery experts, vendors, and end-users from more than 250 organizations have worked together to develop standards and frameworks for addressing e-discovery challenges."[81] By supplying guidelines, standards, whitepapers, research materials, webinars, news, data sheets and other items,[82] the EDRM's model has become "widely accepted and employed by most e-discovery specialists."[83]

EDRM's nine-step flowchart is a conceptual, non-linear, and iterative model of the e-discovery process.[84] The steps include: information management, identification, preservation, collection, processing, review, analysis, production, and presentation.[85] Each step works toward the ultimate goal of translating an excessive volume of documents into relevant and usable material in litigation.[86] The EDRM flow chart "illustrates how the volume of data decreases and the relevance increases as the work progresses."[87] The three steps of processing, reviewing, and analyzing are performed concurrently.[88] While the initial steps of culling data and the final steps of incorporating those materials in a coherent way in litigation are important and present unique challenges in and of themselves, the middle three steps tend to be the areas in which the problems of the data explosion are most often felt and dealt with.[89]

The goal of the processing step is to reduce the volume of ESI and convert it, if necessary, to forms more suitable for review and analysis.[90]

---

79 *See, e.g.,* In re Aspartame Antitrust Litig., 817 F.Supp.2d 608, 614-615 (E.D. Pa. 2011) (in deciding whether to assign the cost of making documents OCR searchable, the court praised the use of OCR in conjunction with keyword searches, which reduced the pool of potentially responsive documents by 87% and 38.5%, for "allowing discovery to be conducted in an efficient and cost-effective manner.").

80 *Frequently Asked Questions*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/joining-edrm/frequently-asked-questions (last visited Dec. 2, 2011).

81 *Id.*

82 *Id.*

83 RALPH C. LOSEY, E-DISCOVERY: CURRENT TRENDS AND CASES 4–5 (2008).

84 *See Electronic Discovery Reference Model Flow Chart, v2.0*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/edrm-stages-explained (last visited Dec. 2, 2011) [hereinafter *EDRM Flow Chart*].

85 *Id.*

86 *See id.*

87 LOSEY, *supra* note 83, at 5.

88 *EDRM Flow Chart*, *supra* note 84.

89 *See* LOSEY, *supra* note 83, at 11–12.

90 *Processing Guide*, ELECTRONIC DISCOVERY REFERENCE MODEL,

To achieve this, practitioners may "reduce the overall set of data collected by filtering out files that are duplicates or known to be irrelevant after further investigation."[91] Duplicate files are removed here,[92] and "[f]iles that are probably not relevant because of factors such as date, type, or origin may also be excluded at this step, if they were not previously excluded" by technicians working during the first four steps.[93] Hot files, or potentially adverse or embarrassing materials, may also be flagged at this stage, as they might have an immediate impact on litigation or make finding similar relevant materials among the remaining files easier.[94]

In processing, practitioners are encouraged to "consider the relationships between the files or documents obtained to better understand what data has been collected and determine whether additional data extraction may be required."[95] The processing step is presented in a linear fashion: moving from assessing to preparing data, to selecting and normalizing, to validating output and exception handling, and then to preparing output and export of the data.[96] Employment of this step is intended to be on an "iterative basis," which means that practitioners often have to make changes to prior tasks and do them again.[97]

The seventh step is the analysis stage.[98] Once relevant materials have been identified, this is the stage where litigation teams attempt to make heads or tails of the information they have, hoping to make informed decisions about strategy and scope through reliable methods based on verified data.[99] Here, litigators identify information such as "key issues, witnesses, specific vocabulary and jargon, and important individual documents."[100] Of course, "[t]his is a traditional legal step that competent trial lawyers are already qualified to perform."[101] Analysis becomes uniquely challenging when large quantities of ESI are produced.[102]

Accordingly, when dealing with large quantities of ESI, the review step becomes the most important and the most difficult. The review step is the point where ESI collected in the previous stages is studied and sorted

---

http://www.edrm.net/resources/guides/edrm-framework-guides/processing (last visited Dec. 2, 2011).

91 LOSEY, *supra* note 83, at 11.

92 *Id.* De-duplication is "[t]he process of identifying (or some vendors include actually removing) additional copies of identical documents in a document collection." *E-Discovery Glossary*, FIOS, http://www.fiosinc.com/e-discovery-knowledge-center/electronic -discovery-glossary.aspx?cid=DG (last visited Dec. 12, 2011).

93 LOSEY, *supra* note 83, at 11.

94 *Id.*

95 *Id.*

96 *See id*.

97 *Id*.

98 *EDRM Flow Chart*, *supra* note 84.

99 *Analysis Guide*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/guides/edrm-framework-guides/analysis (last visited Dec. 12, 2011).

100 LOSEY, *supra* note 83, at 12.

101 *Id.*

102 *Id.*

for use in the latter stages.[103] Here, practitioners "review for relevance, confidentiality and privilege, and related activities such as redaction."[104]

Litigation is made more difficult today by the gigantic hurdles that must be overcome in the document review stage.[105] As practitioners become consumed by this process and expend copious resources to identify usable materials, review becomes an end, rather than a means, to arrive at a legal solution that all parties can agree upon.[106] However, the EDRM model provides a good starting point and has at least alleviated some of the problems caused by large data sets.[107]

### B. Legacy Search and Identification

Although the use of technology in the search and identification phase is not mandated by any court rules, technology is practically required to reduce the amount of manual effort, time, and expense involved in searching for and identifying potentially relevant ESI. Legacy search[108] and identification techniques represent some of the first attempts to harness technology in order to manage large sets of ESI, and are the most widely used today.[109]

### C. Keyword Search Models

Keyword and Boolean search methods are widely used and vetted techniques for filtering data in order to produce responsive documents in discovery.[110] This has much to do with the legal profession's longtime familiarity with major internet legal retrieval services that allow for searches of databases containing statutes and case precedents.[111] However, as recent cases and studies have shown, there are pitfalls to using this

---

103 *Id.*

104 *Id.*

105 *See id.* at 32–33.

106 *See id.*

107 *Review Guide*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/resources/guides/edrm-framework-guides/review (last visited Dec. 12, 2011).

108 *Sedona Conference Glossary: E-Discovery & Digital Information Management,* SEDONA CONFERENCE WORKING GRP. SERIES & WGS MEMBERSHIP PROGRAM, 30 (2007), http://www.thesedonaconference.org/dltForm?did=glossary2010.pdf [hereinafter *Sedona Conference Glossary*] (defining a legacy search as a search of a legacy system, which is "ESI in which an organization may have invested significant resources, but has been created or stored by the use of software and/or hardware that has become obsolete or replaced").

109 *See Best Practices Commentary*, *supra* note 6, at 200.

110 *See* In re CV Therapeutics, Inc. Sec. Litig., No. C-03-3709 SI, 2006 WL 2458720, at *2 (N.D. Cal. Aug. 22, 2006) (endorsing search terms to aid in narrowing production); Windy City Innovations, LLC v. Am. Online, Inc., No. 04 C 4240, 2006 WL 2224057, at *3 (N.D. Ill. July 31, 2006) (promoting keyword searching as a means to search a document more efficiently); In re Lorazepam & Clorazepate Antitrust Litig., 300 F. Supp. 2d 43, 47 (D.D.C. 2004) ("The glory of electronic information is not merely that it saves space but that it permits the computer to search for words or 'strings' of text in seconds."); Medtronic Sofamor Danek, Inc. v. Michelson, 229 F.R.D. 550, 559 (W.D. Tenn. 2003) (ordering defendant to use search terms propounded by plaintiff).

111 *See* Paul & Baron, *supra* note 3, at 21–22.

technique as it often fails to uncover a large portion of potentially relevant data.[112]

A keyword search, in its simplest form, searches for documents possessing a specific term specified by a user.[113] Keyword searches are most often used to identify documents that are either responsive or privileged, and for large-scale culling and filtering of documents.[114] There are limitations with basic keyword searches, however, as they can fail to uncover variants of a word and will not find documents with typographical errors or misspelled words in either the document or query.[115] To address some of the limitations of keyword searches, many databases allow for the use of "wildcards"[116] that enable a user to search for different forms of a certain word.[117]

Boolean searches[118] add another dimension to keyword searches, allowing users to search for multiple keywords together,[119] or exclusive of each other,[120] or within a certain distance from each other.[121] This method "allows multiple keywords or search terms to be linked together to improve the relevancy of the documents identified by this methodology."[122] Other

---

112 *See Best Practices Commentary, supra* note 6, at 200.

113 *The EDRM Glossary*, ELECTRONIC DISCOVERY REFERENCE MODEL, http://www.edrm.net/wiki2/index.php/keyword_search#endnote_vinsonglossary (last visited Oct. 31, 2011); *see also Kroll Ontrack Legal Glossary*, KROLL ONTRACK, http://www.krollontrack.com/glossaryterms (last visited Oct. 31, 2011).

114 *See* Gregory L. Fordham, *Using Keyword Search Terms in E-Discovery and How They Relate to Issues of Responsiveness, Privilege, Evidence Standards, and Rube Goldberg*, 15 RICH. J.L. & TECH. 1, 13, *available at* http://law.richmond.edu/jolt/v15i3/article8.pdf.

115 *See Autonomy's Technology: Limitations of Other Approaches*, AUTONOMY CORP., http://www.autonomy.com/content/Technology/autonomys-technology-limitations-of-other-approaches/index.en.html (last visited Oct. 31, 2011).

116 *See Sedona Conference Glossary*, *supra* note 108, at 54 (defining a wildcard operator as "[a] character used in keyword searching that assumes the value of any alphanumeric character and permits more options, such as alternative spellings, to be identified quickly").

117 KAREN SCHULER ET AL., E-DISCOVERY: CREATING AND MANAGING AN ENTERPRISEWIDE PROGRAM: A TECHNICAL GUIDE TO DIGITAL INVESTIGATION AND LITIGATION SUPPORT 217 (Cathleen P. Peterson & Eva Vincze eds., 2009). The typical wildcard symbol is represented either by an asterisk or an exclamation mark. *Id.* For example, in a sexual harassment case, a search for documents and e-mails containing the wildcard "sex*" or "sex!" might reveal related words such as "sex," "sexual," "sexist," and "sexism." *Id.*

118 "Boolean" refers to the system of logic developed by computer mathematician George Boole. *Best Practices Commentary*, *supra* note 6, at 217.

119 Using the Boolean operator "AND" between two keywords or phrases results in a search that "specifies that both of the items be present for the expression to match." *Search Methodologies*, *supra* note 7.

120 The Boolean operator "OR" used between two keywords or phrases "specifies that either of the two items be present for the expression to match." *Id.*

121 The Boolean operator "W/N" "connects keywords and/or phrases by using a nearness or proximity specification. The specification states that the two words and/or phrases are within *n* words of each other, and the two words/phrases can be in either order." *Id.*

122 *See* SCHULER ET AL., *supra* note 117.

operators include fuzzy searching which can find misspelled terms,[123] and stemming, which search for variations on word endings.[124]

D. E-mail or Conversation Threading

The goal of e-mail or conversation threading is "to find and organize messages that should be grouped together based on reply and forwarding relationships."[125] Typically, an e-mail thread will link together a series of e-mail responses and/or forwards that are created from an original message.[126] This technique may be useful if a particular topic is potentially relevant because responses or forwards of the original message may also contain relevant data.[127] This method is limited, however, when a response to a message changes the subject heading or adds additional information.[128]

E. Shortcomings of Legacy Searches and the Need for Alternatives

Practitioners have adopted legacy search methodologies in earnest, particularly keyword and Boolean searches, to address the expanding universe of information and its associated problems. Courts accept the use of keyword searching to "define discovery parameters and resolve discovery disputes."[129] Despite the widespread use of these techniques, like manual review, keyword searches can be surprisingly inaccurate.[130]

As noted above, the Blair and Maron study revealed a significant gap or disconnect between lawyers' perceptions of their ability to ferret out relevant documents and their actual ability to do so.[131] New research reaffirms the findings of Blair and Maron as applied to keyword searches.[132] In one such study conducted by the Text Retrieval Conference (TREC), researchers found that Boolean keyword searches could only locate between 24% and 57% of the total number of relevant documents.[133] Additionally, these searches produce many false positives, and it is not

---

123 *Id.* at 218.

124 Stemming is a Boolean specification that will "match all morphological inflections of the word." *See Search Methodologies*, *supra* note 7.

125 Sachindra Joshi et al., *Auto-Grouping Emails for Faster E-Discovery*, 4 PVLDB 1284, 1288 (2011), *available at* http://www.vldb.org/pvldb/vol4/p1284-joshi.pdf.

126 *See id.*

127 *See* SCHULER ET AL., *supra* note 117, at 220.

128 *Id.*

129 *See Best Practices Commentary*, *supra* note 6, at 200; *accord* Zubulake v. UBS Warburg LLC, 229 F.R.D. 422, 432 (S.D.N.Y. 2004) (suggesting that a party might satisfy its duty to preserve documents in anticipation of litigation by conducting system-wide keyword searching and preserving a copy of each "hit").

130 *See Best Practices Commentary, supra* note 6, at 194.

131 *See* discussion *supra* Part II.

132 *See, e.g.*, Grossman & Cormack, *supra* note 12, at 18–20; *see also 2010 Legal Track Results*, *supra* note 66; *see also* Roitblat et al., *supra* note 6, at 72.

133 *See* Douglas W. Oard et al., *Overview of the TREC 2008 Legal Track*, *in* NIST SPECIAL PUBLICATION: SP 500-277, THE SEVENTEENTH TEXT RETRIEVAL CONFERENCE (TREC 2008) PROCEEDINGS 8–9 (2008), *available at http*://trec.nist.gov/pubs/trec17/papers/LEGAL.OVERVIEW08.pdf.

uncommon for a poorly chosen keyword to return more "junk" than responsive documents.[134]

Not only are these search methodologies inaccurate, but the adversarial manner by which attorneys use them increases the likelihood that the search will fall short of its target.[135] In an interesting analogy, the method by which most attorneys employ legacy search techniques is similar to the children's game of "Go Fish."[136] When a party requests ESI, the responding party is entitled to privacy and does not have to grant unfettered access to its document database.[137] Yet at the same time, the requesting party is able to make requests for production without revealing what it is that they are looking for.[138] Absent cooperation, the requesting party guesses which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party's "cards," or the terminology used by the responding party's custodians.[139] "This process involves as much chance as skill," takes too long, produces a vast quantity of false positives, and misses many relevant documents.[140]

## III. ALTERNATIVE SEARCH TECHNOLOGIES

Cognizant of the fact that manual review is unworkable and that legacy search methodologies are broken, the Sedona Conference acknowledged that "[t]he legal profession is at a crossroads: the choice is between continuing to conduct discovery as it has 'always been practiced' . . . or, alternatively, embracing new ways of thinking in today's digital world."[141] Indeed, lawyers are gradually beginning to use alternative forms of review with promising results.[142] At the same time, studies demonstrate that these methods, such as concept searching and predictive coding, are able to achieve increasingly higher levels of recall and precision.[143] Moreover, courts are beginning to take notice of the potential

---

134 *Id.* at 7.

135 *See, e.g.,* Ralph Losey, *Child's Game of "Go Fish" Is a Poor Model for E-Discovery Search*, e-Discovery Team Blog (Oct. 4, 2009), http://e-discoveryteam.com/2009/10/04/childs-game-of-go-fish-is-a-poor-model-for-e-discovery-search (suggesting that iterative exchanges of key words and search terms results in poor precision and "a vast quantity of false hits.").

136 *Id.*

137 *See*, *e.g.*, Omnicare, Inc. v. Mariner Health Care Mgmt. Co., No. 3087-VCN, 2009 WL 1515609, at *3 (Del. Ch. May 29, 2009) ("Document discovery must be limited in scope to the production of documents relevant to the subject matter of the litigation between the parties."); Frank v. Engle, No. C.A. 13323, 1998 WL 155553, at *3 (Del. Ch. Mar. 30, 1998) *order clarified sub nom. Lee v. Engle*, No. C.A. 13323-NC 1998 WL 409163 (Del. Ch. June 19, 1998); *see also* JOHN M. BARKETT, E-DISCOVERY: TWENTY QUESTIONS AND ANSWERS 13, 71–77 (2008).

138 Losey, *supra* note 135.

139 *See id.*

140 *Id.* (emphasis omitted).

141 *The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, 10 SEDONA CONF. J. 299, 302 (2009).

142 *See infra* text accompanying notes 163–166.

143 *See generally* Bruce Hedin et al., *Overview of the TREC 2009 Legal Track*, *in* NIST SPECIAL PUBLICATION: SP 500-278, THE EIGHTEENTH TEXT RETRIEVAL CONFERENCE (TREC 2009) PROCEEDINGS (2009), *available at* http://trec.nist.gov/pubs/trec18/papers/

these new methods have to offer.[144] While technology is the source of many of the problems with e-discovery today, technology also represents the solution.[145]

A. Concept Searching

Concept searching allows users to "specify a concept and documents that describe that concept to be returned as the search results."[146] This technique examines the context in which a term appears and looks for similar terms or concepts—a method that is particularly useful in identifying "potentially relevant documents when a set of keywords are not known in advance."[147] When conducted in tandem with legacy search methods such as keyword and Boolean searches, the chance of finding relevant ESI greatly increases.[148]

Concept searches have gained the attention of the courts as well as seen in *Disability Rights Council of Greater Washington v. Washington Metropolitan Transit Authority* (WMATA), a case involving a claim by disabled persons that the WMATA violated the Americans with Disabilities Act and other federal laws.[149] WMATA used an e-mail program that automatically deleted all non-archived e-mail messages every sixty days, and it failed to suspend the deletion program until more than two years after the original complaint was filed.[150] Plaintiffs sought restoration and review of backup tapes to find relevant deleted messages, but WMATA objected, arguing that the backup tapes were not reasonably accessible.[151] The court, however, found support for the plaintiffs' request, determining that the benefit of production outweighed the burden to WMATA,[152] and subsequently ordered the restoration and search of the backups according to a protocol that the parties were directed to negotiate.[153] In doing so, the court suggested that the parties consider using concept searching as opposed to other methods.[154]

---

LEGAL09.OVERVIEW.pdf.

144 *See infra* text accompanying notes 149–154.

145 *See* Hon. Andrew J. Peck & David J. Lender, *10 Key E-Discovery Issues in 2011: Expert Insight to Manage Successfully*, METRO. CORP. COUNSEL, Apr. 2011, at 5, *available at* http://www.metrocorpcounsel.com/pdf/2011/April/01.pdf ("Since technology created the volume, lawyers have turned to technology to attempt to solve the review problem . . . .").

146 *Search Methodologies*, *supra* note 7.

147 *Id.*

148 SCHULER ET AL., *supra* note 117, at 220.

149 *See, e.g.*, Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139 (D.D.C. 2007).

150 *Id.* at 145.

151 *Id.* at 147.

152 *Id.* at 148 (applying Rule 26(b)(2)(C)'s balancing test, the magistrate judge determined that there was no other place to find the documents due to WMATA's failure to impose a litigation hold, that the discovery was important to the outcome of the litigation, and the plaintiffs had no meaningful financial resources).

153 *Id.*

154 *See id.* (Magistrate Judge Facciola questioned the search methods of the restored data as follows:

B. Predictive Coding

Another automated search method that has recently gained attention is predictive coding, or computer-assisted coding.[155] These coded documents are then used by the computer system in an iterative process to code additional documents across the full collection.[156] This process merely accelerates the discovery process; it does not replace manual review by humans, but optimizes it.[157]

The reviewing human typically codes a controlled sample group of documents based on a series of "yes" or "no" questions, such as whether each document is responsive, relevant, or privileged.[158] "The system builds an ontology in the background as it learns from the expert and presents subsequent samples."[159] After running enough iterations, the system will have "sufficiently built the ontology to the point where it can 'predict' what the human will" pick out in the sample he or she is reviewing.[160] Considering that manual review can be effective in small samples,[161] predictive coding efficiently combines a human's analytical assessments with the processing power of a computer.[162]

As mentioned above, the results of the TREC 2008 Legal Interactive Task study suggest that predictive coding may in fact be able to improve upon manual review and legacy search methods.[163] One participant in the study employed predictive coding in response to a mock request to produce documents from a collection of 6,910,192 documents.[164] By coding a

---

> how will they be searched to reduce the electronically stored information to information that is potentially relevant? In this context, I bring to the parties' attention recent scholarship that argues that concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results.)

(citing Paul & Baron, *supra* note 3).

155 *See, e.g.*, Baron & Berman, *supra* note 51, at 7 (citing *E-Discovery Institute Survey on Predictive Coding*, e-DiscoveryInst. (2010), http://www.ediscoveryinstitute.org/publications/ediscovery_institute_survey_on_predictive_coding (describing predictive coding as

> a combination of technologies and processes in which decisions pertaining to the responsiveness of records gathered or preserved for potential production purposes . . . are made by having reviewers examine a subset of the collection and having the decisions on those documents propagated to the rest of the collection without reviewers examining each record.

156 *Id.*

157 *See id.*

158 Tom Groom, *Applying Predictive Coding to Reduce Costs and Increase Quality in Document Review*, D4 DISCOVERY, http://www.d4discovery.com/2012/02/applying-predictive-coding-to-reduce-costs-and-increase-quality-in-document-review/ (last visited Oct. 3, 2011).

159 *Id.* An ontology is "[a] collection of categories and their relationships to other categories and to words." *Sedona Conference Glossary*, *supra* note 108, at 37.

160 Groom, *supra* note 158.

161 *See generally* Blair & Maron, *supra* note 67.

162 *See* Joe Dysart, *A New View of Review: Predictive Coding Vows to Cut E-Discovery Drudgery*, ABA J., Oct. 2011, at 26.

163 *See generally* Oard et al., *supra* note 133.

164 *See* Grossman and Cormack, *supra* note 12, at 19-20 (citing Christopher Hogan et al., *H5 at TREC 2008 Legal Interactive: User Modeling, Assessment & Measurement*, *available at* http://trec.nist.gov/pubs/trec17/papers/h5.legal.rev.pdf.

smaller sample of documents and inputting them into the computer, the researchers examined only 7,992 documents, approximately 860 times fewer than would have been necessary to complete an exhaustive manual review.[165] Still, the results compared favorably to the other search methods, as the researchers achieved recall rates ranging between 62.4% and 81.0%, far exceeding the 20.26% average recall rate in the Blair and Maron study.[166]

## IV. RECENT CASE LAW ON REASONABLE SEARCH PROTOCOLS

Courts have yet to embrace any of the new search technologies, instead only generally alluding to potential benefits they offer, but not going so far as to expressly endorse a particular method. In the meantime, lawyers must still meet their discovery obligations and defend the decisions they made when challenged on their selection of relevant materials. While keyword searching may be the most widely available and employed option, it is still quite possible to use an inadequate Boolean search.[167]

Until only recently, few cases offered guidance on the reasonableness of electronic searches in e-discovery. In 2006, Congress passed a set of amendments to the Federal Rules of Civil Procedure,[168] sometimes known as the "ESI Amendments," however these did not mention the use of electronic searches.[169] Decisions regarding manual review only offered nominal guidance since they did not address the technological complexities of electronic searches.[170]

Before 2007, the Sedona Conference's *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery* was

---

165 *Id.*

166 *Id.*; *see generally* Blair & Maron, *supra* note 70.

167 *See* Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 256–57 (D. Md. 2008). In a case where a keyword search was designed to locate ESI and not privileged materials, the court held that by voluntarily, yet inadvertently, producing a series of ESI to the opposing party, the party waived attorney-client privilege and work product protection for the documents. *Id.* at 253–54. The court noted "while it is universally acknowledged that keyword searches are useful tools for search and retrieval of ESI, all keyword searches are not created equal; and there is a growing body of literature that highlights the risks associated with conducting an unreliable or inadequate keyword search or relying exclusively on such searches for privilege review." *Id.* at 256–57; *see also, e.g.,* ClearOne Commc'ns, Inc. v. Chiang, 2008 WL 920336, at *5–7 (D. Utah, Apr. 1, 2008).

168 *See* FED. R. CIV. P. 16, 26, 33, 34, & 45 (2006).

169 *Id.*

170 *See* FED. R. CIV. P. 26(b)(2) (2006). The Rule dictates a two-tiered approach to the production of ESI, only making a distinction between that which is reasonably accessible and that which is not. *Id.* The Advisory Committee notes that "[t]he information explosion of recent decades has greatly increased both the potential cost of wide-ranging discovery and the potential for discovery to be used as an instrument for delay or oppression," and intended Rule 26(b)(2) "to provide the court with broader discretion to impose additional restrictions on the scope and extent of discovery." FED. R. CIV. P. 26 (Advisory Committee's note on 1993 amendments); *see also* Am. Int'l Specialty Lines Ins. Co. v. NWI-I, Inc., 240 F.R.D. 401, 412 (N.D. Ill. 2007) (holding that a party would not be required to review 19,000 boxes of documents where both the issues and the resources of the parties were limited); *cf.* Alexander v. FBI, 194 F.R.D. 305, 315–16 (D.D.C. 2000) (finding that a search of indices of about sixty boxes of documents, rather than reviewing every document, was inadequate).

the only significant resource concerning the reasonableness of e-discovery search methods.[171] The commentary's goal was to provide a guide on the "nature of the search and retrieval process."[172] However, while the commentary discusses keyword searches as a useful method to find particular documents, it notes its shortcomings in certain contexts and does not suggest a particular alternative search method.[173]

More recently, however, a few opinions have attempted to provide guidance on what methods constitute reasonable searches. *United States v. O'Keefe*, *Equity Analytics, LLC v. Lundin*, and *Victor Stanley, Inc. v. Creative Pipe, Inc.* all suggest that keyword searching may not be sufficient.[174] Moreover, *Victor Stanley* goes on to trumpet alternative search methods in certain circumstances, but does not expressly endorse a preferred technique.[175] Common throughout all three cases is the requirement that attorneys be prepared to defend their search methods if challenged, and that such preparation may involve the use of a technical expert, or at least someone with the qualifications needed to design and implement an effective search methodology.[176]

### A. *United States v. O'Keefe*

*O'Keefe* suggests that expert evidence may be required to evaluate the efficacy of a keyword search in identifying responsive documents.[177] In *O'Keefe*, the court found a number of inadequacies in the government's search for records and concluded that it had failed to comply with a discovery order.[178] Despite this, the court rejected the defendants' argument regarding the adequacy of the search terms used by the government, holding that the defendants would have had to specifically contend that the search terms used by the government were insufficient in a separate motion to compel, which would be based on evidence rising up to the requirements of Rule 702 of the Federal Rules of Evidence.[179]

---

171 *See Best Practices Commentary*, *supra* note 6.

172 *Id.* at 191.

173 *Id.* at 201–04; *see also supra* text accompanying notes 129–140.

174 U.S. v. O'Keefe, 537 F. Supp. 2d 14, 23–24 (D.D.C. 2008); Equity Analytics v. Lundin, 248 F.R.D. 331, 332–33 (D.D.C. 2008); Victor Stanley v. Creative Pipe, Inc., 250 F.R.D. 251, 260 (D. Md 2008).

175 *Victor Stanley*, 250 F.R.D. at 259 n.9 (noting that electronic search methods "can enhance the accuracy and reliability of the search.").

176 *See infra* Part IV, A, B, and C.

177 *O'Keefe*, 537 F. Supp. 2d at 24.

178 *Id.* at 17–22.

179 *Id.* at 24; FED. R. EVID. 702 (requiring that for an expert to testify,

> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.);

*see also* FED. R. EVID. 702 advisory committee's note (2000) (noting that Rule 702 provides "general standards that the trial court must use to assess the reliability and helpfulness of proffered expert testimony.").

Defendant O'Keefe, a Department of State employee, was indicted for allegedly receiving gifts from co-defendant Agrawal in return for expediting visa requests for employees of Agrawal's company.[180] Whether such requests were expedited routinely by various consulates without receipt of anything of value became an issue, and the court ordered the government to search both its hard copy and electronic files for responsive documents.[181]

After receiving the government's production, the defendants filed a motion to compel, protesting that the government had not met the judge's order.[182] The defendants expressed concern that the government had not had its employees search their own electronically stored information for documents, making it "impossible to identify the source or custodian of [each] document."[183] Moreover, they contended that the government had not revealed what steps it had taken to preserve documents.[184]

The court concluded that the defendants' concern over deficiencies in the government's production of electronically stored information was "an insufficient premise for judicial action."[185] By analogy, Rule 37(e) of the Federal Rules of Civil Procedure provided that sanctions were inappropriate if loss of such information was the result of the "routine, good-faith operation of an electronic information system."[186] Thus, if the defendants intended to charge that the government destroyed evidence that should have been preserved, the claim would have to be based on direct evidence.[187] It would not be enough to surmise that they should have received more than they did.[188]

The court also held that any contention of the defendants that search terms used by the government were ineffective would have to be made in a motion to compel supported with expert testimony pursuant to Rule 702 of the Federal Rules of Evidence.[189] The sufficiency of search terms was "a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics."[190] The court also cited the Sedona Conference's *Best Practices Commentary*[191] and noted the limitations of keyword searches, but went on to explain that evaluating particular search methodologies is not easy:

---

180 *O'Keefe*, 537 F. Supp. 2d at 15–16.
181 *Id.* at 16.
182 *Id.*
183 *Id.* at 18.
184 *Id.* at 22–23.
185 *Id.* at 20 (referencing Hubbard v. Potter, 247 F.R.D. 27, 30–31 (D.D.C. 2008)).
186 *Id.*
187 *Id.* at 22–23.
188 *Id.*
189 *Id.* at 24.
190 *Id.* (referencing Paul & Baron, *supra* note 3).
191 *Id.* (referencing *Best Practices Commentary*, *supra* note 6).

> [F]or 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. This topic is clearly beyond the ken of a layman and requires that any such conclusion be based on evidence that, for example, meets the criteria of Rule 702 of the Federal Rules of Evidence.[192]

Thus, the court declined to address the question of the reliability of the search method without a motion to compel, supported by expert testimony under Rule 702 of the Federal Rules of Evidence.[193]

### B. *Equity Analytics, LLC v. Lundin*

Similarly, the *Equity Analytics* court suggests that Rule 702 expert evidence may be required to evaluate the methods employed to collect documents.[194] The court in *Equity Analytics* stated that determining whether "a particular search methodology, such as keywords, will or will not be effective certainly" requires "knowledge beyond the ken of a lay person (and a lay lawyer) and requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence."[195]

In *Equity Analytics*, the court was asked to resolve the dispute between the parties in their attempt to develop a search protocol for examination of the defendant's computer.[196] The plaintiff alleged that the defendant had gained illegal access to the plaintiff's electronically stored information after the defendant was fired by the plaintiff.[197] The defendant's computer contained a wide range of materials, many having nothing to do with the lawsuit.[198] The defendant opposed production of the data and proposed that only certain file types be searched but the plaintiff objected.[199]

The court declined to determine whether the proposed search was adequate based on the arguments of the attorneys alone, instead requiring the plaintiff to submit an affidavit from an expert explaining why the narrow search proposed by the defendant was not enough.[200] The court reasoned that such expert testimony would provide the information needed to best assess how to balance the plaintiff's need for information with the privacy of the defendant.[201]

These two cases suggest that expert evidence may be required to assess the searches, and that experts may be needed prior to and during

---

192 *Id.*
193 *Id.*
194 Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008).
195 *Id.*
196 *Id.* at 332.
197 *Id.* at 331–32.
198 *Id.* at 332.
199 *Id.* at 332–33.
200 *Id.* at 333.
201 *Id.*

litigation to design search techniques to ensure that the searches will be defensible.

### C. *Victor Stanley, Inc. v. Creative Pipe, Inc.*

Like *O'Keefe* and *Equity Analytics*, *Victor Stanley* requires qualified persons to craft an effective search methodology, but it does not go so far as to require an expert.[202] However, through discussion of electronic searches, the court offers a more practical standard for assessing search protocols.[203]

The issue in *Victor Stanley* was whether the defendants had waived attorney-client privilege for documents that counsel had accidentally produced.[204] The defendants had used keyword searches to identify non-privileged documents.[205] One of the defendants, and two of the defendants' lawyers, chose about seventy keywords for their expert to use in searching for protected documents in the defendants' ESI through use of a search protocol agreed upon with the plaintiff.[206] They did not, however, manually review any of the results of that search for privilege.[207] In addition, the defendants did a manual privilege review of the titles of some documents that were reportedly not text-searchable.[208] Despite the expert's search for protected documents, the plaintiff alerted the defendants to documents in the production that appeared to be privileged or protected.[209] The defendants sought the return of these documents, but the plaintiff countered that the defendant had waived privilege.[210]

Since the case was decided before Rule 502 of the Federal Rules of Evidence was adopted,[211] the court used a five-factor test from *McCafferty's, Inc. v. Bank of Glen Burnie* to evaluate whether the defendant had waived privilege.[212] The factors in *McCafferty's* include: "(1) the reasonableness of the precautions taken to prevent inadvertent disclosure, (2) the number of inadvertent disclosures, (3) the extent of the disclosure, (4) any delay in measures taken to rectify the disclosure, and (5) overriding interests in justice."[213] Of particular note in *Victor Stanley* is the reasonableness factor, which is similar to the requirement in Rule 502(b)(2)

---

202 *See* Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260–61 n.10 (D. Md. 2008).

203 *Id.* at 259–61 nn. 9 & 10.

204 *Id.* at 253, 255.

205 *Id.* at 256.

206 *Id*. At 255–56.

207 *Id.* at 257 ("[I]t appears from the information that they provided to the court that they simply turned over to the Plaintiff all the text-searchable ESI files that were identified by the keyword search Turner performed as non-privileged . . . .").

208 *Id.* at 256–57.

209 *Id.* at 255.

210 *Id.* at 255, 257.

211 *See* FED. R. EVID. 502 (2008) (limiting the consequences of waiver when privileged documents are disclosed).

212 *Victor Stanley*, 250 F.R.D. at 259.

213 McCafferty's, Inc. v. Bank of Glen Burnie, 179 F.R.D. 163, 167–68 n.9 (D. Md. 1998).

of the Federal Rules of Evidence, which requires analysis of whether "the holder of the privilege or protection took reasonable steps to prevent disclosure" in assessing whether a disclosure results in waiver.[214]

*Victor Stanley* went on to hold that the defendants had waived privilege, finding they had failed to meet their burden to establish that their search was satisfactory because of their failure to identify the keywords they used to conduct the searches, to explain why they chose the keywords, and to explain what type of search was done.[215] The court spent considerable space discussing this latter failure, stating that "for the benefit of future cases," parties should state the procedures they follow in the process of conducting searches, and the court then further provided a lengthy footnote summarizing search methodologies discussed in the Sedona Conference's Best Practices Commentary.[216]

One aspect of the defendants' failure of proof was that they did not show how the defendants and their attorneys were qualified to design the search that they used and analyze the results of the search to assess its reliability, appropriateness, and implementation.[217] The court observed that when it comes to keyword searches, the "proper selection and implementation obviously involves technical, if not scientific knowledge."[218] *Victor Stanley* does not go as far as *O'Keefe* and *Equity Analytics* in suggesting that the person who makes a search protocol must be an expert under Rule 702 of the Federal Rules of Evidence, but *Victor Stanley* nevertheless holds that for a contested search to withstand judicial scrutiny, a party must be able to justify the steps it undertook.

## V. MOVING FORWARD

E-discovery decisions should always be based on honoring the goal of Rule 1 of the Federal Rules of Civil Procedure: "the just, speedy, and inexpensive determination of every action and proceeding."[219] One way to accomplish this is by adopting advanced search methodologies. While advanced search techniques are becoming more ubiquitous, progress remains slow.

> Litigators may accept simple keyword searching, yet be reluctant to use alternative search techniques. They may not be convinced that the chosen method would withstand a court challenge. They may perceive a risk that problem documents will not be found despite the additional effort; and an opposite risk that

---

214 *See* FED. R. EVID. 502(b)(2) (2011).

215 *Victor Stanley*, 250 F.R.D. at 258–59

216 *Id.* at 259 n.9, 264; *see also supra* Part III.

217 *Victor Stanley*, 250 F.R.D. at 259–60 (citing U.S. v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008); Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008)).

218 *Victor Stanley*, 250 F.R.D. at 260.

219 *See* FED. R. CIV. P. 1 (2007); *The Sedona Conference Cooperation Proclamation*, 10 SEDONA CONF. J. 331, 333 [hereinafter *Cooperation Proclamation*].

> documents might be missed which would otherwise be picked up in a straight keyword search.[220]

Compounding this problem, however, is the lack of express judicial approval for these search technologies.[221] For example, to date, no reported case, federal or state, has ruled on the use of predictive coding. It is possible that many attorneys are reluctant to act as the proverbial "guinea pig[s]," waiting for official guidance on how to proceed in these types of searches first.[222] Magistrate Judge Andrew Peck pondered this issue in a recent editorial, offering the following:

> Perhaps they are looking for an opinion concluding that: "It is the opinion of this court that the use of predictive coding is a proper and acceptable means of conducting searches under the Federal Rules of Civil Procedure, and furthermore that the software provided for this purpose by [insert name of your favorite vendor] is the software of choice in this court." If so, it will be a long wait.[223]

Aside from possible breaches of judicial ethical rules, there are presumably various reasons for this. As the Sedona Conference's Practice Point 3 states, "[t]he choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed."[224] Formal support for a particular search technique is an impractical one-size-fits-all approach[225] that ignores variables that change from case to case, including how the search application was used, by whom, the type of case, alternatives that were or should have been considered, and cost.[226]

This is not to say that the bench does not support the use of innovative search methodologies in discovery; in fact, the reality is quite the opposite. Judge Peck himself expresses support for judicial decisions critiquing keyword searches, particularly *O'Keefe*, *Equity Analytics*, and *Victor Stanley*.[227]

In *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.*, Judge Peck notably issues "a wake-up call to the Bar in this District about the need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or 'keywords' to be used to produce emails or other electronically

---

220 *Best Practices Commentary*, *supra* note 6, at 203.

221 *See id.*

222 Hon. Andrew Peck, *Search, Forward: Will Manual Document Review and Keyword Searches be Replaced by Computer-Assisted Coding?*, L. TECH. NEWS (Oct. 1, 2011).

223 *Id.*

224 *Best Practices Commentary*, *supra* note 6, at 194.

225 *See* Matthew Prewitt, *E-Discovery: One Size Does Not Fit All*, INSIDE COUNSEL (Sept. 20, 2011), http://www.insidecounsel.com/2011/09/20/e-discovery-one-size-does-not-fit-all.

226 Chris Dale, *Judge Peck and Predictive Coding at the Carmel E-Discovery Retreat*, E-DISCLOSURE INFO. PROJ. (Aug. 2, 2011), http://chrisdale.wordpress.com/2011/08/02/judge-peck-and-predictive-coding-at-the-carmel-ediscovery-retreat.

227 *See* Peck, *supra* note 222.

stored information."[228] The problem in *William Gross* was not the keyword technology that was used, but the failure of the parties to come to an agreement on a list of keywords.[229] When the responding party deployed overbroad and imprecise keyword search terms to respond to a discovery request, Judge Peck bemoaned the case as "the latest example of lawyers designing keyword searches in the dark, by the seat of the pants, without adequate . . . discussion with those who wrote the emails."[230]

The court ordered a multi-step framework that the litigators must use when selecting a keyword search strategy.[231] Judge Peck ordered that the attorneys "at a minimum must carefully craft the appropriate keywords, with input from the ESI's custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of 'false positives.'"[232]

Judge Peck's opinion in *William Gross* demonstrates how courts are developing factors to assess the reasonableness of a litigant's search methodology on a case-by-case basis rather than assessing search methodologies individually and out of context. The two most important factors are "cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI."[233]

## VI. COOPERATION IS KEY

Cooperation is being touted by an increasing number of courts as an effective way to reduce the costs and risks of e-discovery.[234] As the court in *Mancia v. Mayflower Textile Services Company* explains, cooperation among counsel "will almost certainly result in having to produce less discovery, at lower cost . . . [and] will almost certainly result in getting helpful information more quickly" for the requesting parties.[235]

Parties should attempt to cooperate with opposing counsel to agree on a discovery plan that sets forth specific protocols for identifying responsive

---

228 William A. Gross Constr. Assoc., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 134 (S.D.N.Y. 2009).

229 *Id.* at 134–35.

230 *Id.*

231 *See id.* at 136.

232 *Id.*

233 *Id.*; *see also Cooperation Proclamation*, *supra* note 219; Part IV, *supra* (discussing the use of e-discovery experts in order to maintain a defensible position in litigation).

234 *See, e.g.,* SEC v. Collins & Aikman Corp., 256 F.R.D. 403, 414 (S.D.N.Y. 2009) (finding the SEC's "refusal to negotiate a workable search protocol . . . 'patently unreasonable'"); Mancia v. Mayflower Textile Servs. Co., 253 F.R.D. 354, 365 (D. Md. 2008) (stating that cooperation is advantageous to both parties). *See generally*, Aguilar v. Immigration and Customs Enforcement Div. of U.S. Dep't of Homeland Sec., 255 F.R.D. 350 (S.D.N.Y.2008); Capitol Records, Inc. v. MP3tunes, LLC, 261 F.R.D. 44, 47-48 (S.D.N.Y. 2009); Ford Motor Co. v. Edgewood Prop., Inc., 257 F.R.D. 418 (D.N.J. 2009); Newman v. Borders, Inc., 257 F.R.D. 1 (D.D.C. 2009); Romero v. Allstate Ins. Co., 271 F.R.D. 96 (E.D. Pa. 2010); Oracle USA, Inc. v. SAP AG, 264 F.R.D. 541 (N.D. Cal. 2009).

235 *Mancia*, 253 F.R.D. at 365.

and privileged documents.[236] Courts are just as quick to reward parties that cooperate as they are to punish those that do not.[237]

In terms of developing search protocols, a party's failure to cooperate can have dramatic effects beyond driving up the cost of litigation[238] and overburdening the justice system.[239] For example, the court in *William A. Gross* did not willingly decide to order its own search protocol—instead, the court opined that the parties' inability to agree put the court "in the uncomfortable position of having to craft a keyword search methodology for the parties, without adequate information from the parties."[240] Moreover, a court may even be motivated to shift discovery costs to uncooperative parties.[241]

Perhaps the benefits of cooperating are best realized if parties are able to work together from the outset, as this may decrease the chance that a dispute about the search efforts taken by each party will later develop.[242] With regard to search terms protocols, parties can further this goal by collaborating on which search process to use, the terms to be used in that process, and by agreeing to participate in an iterative process where successive searches can be modified and improved upon.[243] On a more fundamental level, parties are encouraged to come to the table armed with knowledge of likely sources of ESI, its custodians, and understanding of the steps and costs required to access the ESI.[244] A party's own preparation in this area can help facilitate cooperation and smooth discovery.[245]

---

236 *See Best Practices Commentary, supra*, note 6, at 211.

237 *See, e.g.*, In re Seroquel Prods. Liab. Litig, 224 F.R.D. 650, 664–65 (M.D. Fla. 2007) (finding sanctions were warranted because party failed to produce reasonably accessible documents); *The Case for Cooperation*, 10 SEDONA CONF. J. 339, 359 (2009 Supp.) (discussing the strategic benefits of cooperation).

238 Mancia v. Mayflower Textile Services. Co., 253 F.R.D. 354, 359 (D. Md. 2008) (noting that "failure to engage in discovery . . . is one reason why the cost of discovery is so widely criticized as being excessive—to the point of pricing litigants out of court.").

239 *The Case for Cooperation*, *supra note* 239, at 343 (describing how a lack of cooperation between parties may "prevent[] adjudication of meritorious claims").

240 William A. Gross Construction Assoc., Inc. v. Am. Mfrs Mutual Ins. Co., 256 F.R.D. 134, 135 (S.D. NY 2009).

241 *See* Surplus Source Grp., LLC v. Mid Am. Engine, Inc., 2009 U.S. Dist. LEXIS 29260, *4-5 (E.D. Tex. 2009) (shifting costs to the plaintiff where although the defendant demonstrated "a persistent willingness to aide [sic] the Plaintiffs in crafting an ESI search," the plaintiff unreasonably delayed in responding to defendant's attempt to negotiate.).

242 "Early agreement . . . makes it much less likely that a party will be ordered to supplement its production . . . because its opponent convinces a court that the producing party's unilateral choices were too narrow or otherwise inappropriate." *The Case for Cooperation*, *supra* note 239, at 358.

243 Steven S. Gensler, *A Bull's-Eye View of Cooperation in Discovery*, 10 Sedona Conf. J. 363, 371 (2009).

244 *The Case for Cooperation*, *supra* note 239, at 344.

245 *See* Covad Commc'ns Co. v. Revonet, Inc., 254 F.R.D. 147, 151 (D.D.C. 2008) (stating that "the courts have reached the limits of their patience with having to resolve electronic discovery controversies that are . . . so easily avoided by the lawyers' conferring with each other on such a fundamental question as the format of their productions of electronically stored information.").

## VII. QUALITY CONTROL OF DEFENSIBLE SEARCH PROTOCOLS

Despite its strong support for advanced electronic search tools, the Sedona Conference notes that "[t]echnologically advanced tools, however, 'cutting edge' they may be, will not yield a successful outcome unless their use is driven by people who understand the circumstances and requirements of the case, as guided by thoughtful and well-defined methods, and unless their results are measured for accuracy."[246] This underscores the importance of strategically planning, documenting, and supervising the entire e-discovery process. Also, as the Sedona Conference makes clear, "parties should expect that their choice of search methodology will need to be explained . . . in subsequent legal contexts (including depositions, evidentiary proceedings, and trials)."[247]

A party should be ready to place its discovery plan's effectiveness on the line by including a method for testing and assessing the effectiveness of their search protocols, and evaluating recall and precision rates either by sampling supposed nonresponsive documents and/or documents reviewed during the primary review phase.[248] If parties wish, they may also employ third party professionals to sample the effectiveness ofa set of search protocols.[249] As discussed above, parties should consider retaining experts to develop, execute, and defend a protocol when appropriate.[250]

Aside from enabling a party to more adequately defend itself should a dispute over discovery arise, such practices promote self-policing and quality control.[251] In doing so, parties will more reliably know at the end of the discovery stage how accurate and complete their methods were, and will not be left to question whether they violated the duty to preserve, uncover, or disclose relevant evidence and the possibility that privilege or confidential information may have been inadvertently produced.[252]

The message to be taken from the cases of *O'Keefe*, *Equity Analytics*, and *Victor Stanley* is clear: when parties decide to use a particular ESI search method, it needs to be aware of the intricacies of its own storage system and craft a discovery plan accordingly. Should an opposing party challenge the method selected, the discovery propoent should then expect to support its position with this information, perhaps with the assistance of discovery experts.

---

246 *The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, *supra* note 141, at 306.

247 *See Best Practices Commentary*, *supra* note 6, at 212.

248 *See The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, *supra* note 141, at 310–12.

249 *Id.* at 310.

250 *See supra* Part IV.

251 *See generally id.* at 309–10.

252 *See generally id.*

## CONCLUSION

The information inflation shows no signs of slowing down. Parties to litigation have the choice of confronting this problem head on by embracing newer and more technologically advanced search methodologies, or proceeding at their own risk as they have before.[253] In doing so, they face rising expenditures of time and money because their search and retrieval method is unlikely to be the most efficient or reliable possibility. Regardless of which method they choose to adopt, parties should unquestionably engage in cooperative efforts to arrive at agreeable search protocols, and develop and document thoughtful discovery plans from the ground up so as to best defend their own discovery practices and decisions.[254]

## POSTSCRIPT/UPDATE TO E-DISCOVERY NOTE

Following completion of the drafting of this Note in December 2011, computer assisted review has steadily gained attention and, for the first time, acceptance in the legal community.[255]

As discussed above, in October 2011, Magistrate Judge Andrew Peck suggested that despite a number of judicial opinions highly critical of keyword searching, one reason many attorneys have been slow to adopt new search technology is that they apparently "are waiting for a judicial decision approving of computer-assisted review. . . . If so, it will be a long wait."[256]

Interestingly, this "long wait" turned out to be just over four months. In February 2012, Judge Peck issued an opinion approving the use of predictive coding.[257] In doing so, Judge Peck specifically noted that his opinion in *Da Silva Moore v. Publicis Groupe* "appears to be the first in which a Court has approved of the use of computer-assisted review."[258]

In his October 2011 periodical, Judge Peck set forth guidelines for handling discovery challenges to any proposed use of computer-assisted review that came before him.[259] In this situation, Judge Peck stated that he would pay close attention to the process and results of the search:

---

253 *Supra* Part I.

254 *Supra* Part IV.

255 *See, e.g.*, Da Silva Moore v. Publicis Groupe, No. 11 Civ. 1279(ALC)(AJP), 2012 WL 607412, at *1 (S.D.N.Y. Feb. 24, 2012) (approving the use of computer assisted review); Order Approving the Use of Predictive Coding for Discovery, Global Aerospace, Inc. v. Landow Aviation, L.P., No. CL 61040, 2012 WL 1431215 (2012), at *1 (ordering defendants to proceed with predictive coding); Nat'l Day Laborer Org. Network v. U.S. Immigration & Customs Enforcement Agency, No. 10 Civ. 3488 (SAS), 2012 WL 2878130, at *12 (S.D.N.Y. July 13, 2012) (discussing the effectiveness of computer assisted coding).

256 *See* Peck, *supra* note 222.

257 *See generally* Da Silva Moore v. Publicis Groupe, No. 11 Civ. 1279(ALC)(AJP), 2012 WL 607412 (S.D.N.Y. Feb. 24, 2012).

258 *Id.* at *12.

259 *See* Peck, *supra* note 222.

> [I]f the use of predictive coding is challenged in a case before me, I will want to know what was done and why that produced defensible results. I may be less interested in the science behind the 'black box' of the vendor's software than in whether it produced responsive documents with reasonably high recall and high precision . . . . That may mean allowing the requesting party to see the documents that were used to train the computer-assisted coding system . . . . Proof of a valid 'process,' including quality control testing, also will be important.[260]

Judge Peck's opinion in *Da Silva Moore*closely mirrored the reasoning he set forth in the periodical. In *Da Silva Moore*, a Title VII action, the plaintiffs objected to defendant MSLGroup's use of predictive coding "to cull down" over three million documents involved in discovery.[261] The parties agreed to use computer-assisted review but disagreed over how it should be implemented, with the plaintiffs claiming that MSL's proposal to use a number of rounds to test and refine the searches and review software, and to share the seed documents and documents flagged as relevant or irrelevant, was not reliable or transparent.[262]

Noting his earlier writings to the contrary, Judge Peck plainly held that "[t]his judicial opinion now recognizes that computer-assisted review is an acceptable way to search for relevant ESI . . . ."[263] However, Judge Peck cautioned that this does not mean computer-assisted review should be used in all cases, or that the exact ESI protocol approved in *Da Silva Moore* will be appropriate in all future cases that utilize computer-assisted review. Rather, he noted "computer-assisted review is not a magic, Staples–Easy–Button, solution appropriate for all cases."[264] While admitting that it is "not perfect," Judge Peck determined that computer-assisted review was better than the alternatives in the case at bar.[265] Judge Peck further encouraged parties to "seriously consider[] [computer-assisted review] for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review."[266]

With *Da Silva Moore* leading the way, a number of other courts have quickly followed suit and have begun to entertain predictive coding as a viable tool in discovery.[267] In the plaintiffs' challenge to *Da Silva Moore*, Judge Andrew Carter approved Judge Peck's ruling and written order supporting computer-assisted review.[268] Furthermore, a Virginia state court approved a computer-assisted review protocol proposed by the defendants

---

260 *Id.*
261 *Moore*, 2012 WL 607412, at *3.
262 *Id.* at *3–8, *1 n.1.
263 *Id.* at *1.
264 *Id.* at *8.
265 *Id.* at *11.
266 *Id.* at *12.
267 *See supra* note 255 and accompanying text.
268 *Moore*, 2012 WL 1446534, at *2.

in their protective order for purposes of processing and producing ESI.[269] Yet another court criticized the shortcomings of keyword searches and endorsed predictive coding to "allow humans to teach computers what documents are and are not responsive to a particular FOIA or discovery request and . . . [to] significantly increase the effectiveness and efficiency of searches."[270] Finally, in discussing a scheduling order from the Delaware Court of Chancery, a judge even instructed the parties, without any outside cues, to adopt a predictive coding strategy or demonstrate good cause to avoid it.[271]

In addition to merely lending judicial legitimacy to computer-assisted review, the trend affirms this Note's emphasis on cooperation between parties and transparency in all aspects of preservation and production of ESI.[272] Citing the Sedona Conference, Judge Peck reiterated in *Da Silva Moore* that "'the best solution in the entire area of electronic discovery is cooperation among counsel.'"[273] One reason why Judge Peck ordered computer-assisted review protocols was that MSL's "transparency allow[ed] the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called 'black box' of the technology."[274] In upholding Judge Peck's order, Judge Carter further trumpeted cooperation and transparency as key ingredients in computer-assisted discovery, stating that since the "ESI protocol . . . builds in levels of participation by Plaintiffs," the plaintiffs will have opportunity to shape the process and thus ensure it meets their needs.[275] Furthermore, in resolving a dispute surrounding a party's interrogatories and document requests, another court channeled the principles of cooperation of the Sedona Conference, urging counsel not to "confuse advocacy with adversarial conduct" in addressing discovery obligations.[276] Additionally, the use of experts is cited as a valuable tool to evaluate the efficacy of a search protocol in furtherance of these efforts.[277]

With support for technology-assisted review gaining momentum among the judiciary, parties can better position themselves to ride the coming wave by cooperating actively with opposing counsel, developing sensible discovery plans and being prepared to defend them, and sharing these protocols openly and transparently as appropriate.

---

269 Order Approving the Use of Predictive Coding for Discovery, 2012 WL 1431215 (2012).

270 *Nat'l Day Laborer Org. Network*, 2012 WL 2878130, at *11–12.

271 *See* EORHB, Inc., et al v. HOA Holdings, LLC, C.A. No. 7409-VCL (Del. Ch. 2012).

272 *See supra* Part VI.

273 *Moore*, 2012 WL 607412, at *11 (quoting *Cooperation Proclamation*, *supra* note 220).

274 *Id.*

275 Moore v. Publicis Groupe SA, No. 11 Civ. 1279(ALC)(AJP), 2012 WL 1446534, at *2 (S.D.N.Y. Apr. 26, 2012).

276 Kleen Prods. LLC v. Packaging Corp. of Am., 2012 U.S. Dist. LEXIS 139632, *6 (N.D. Ill. 2012) (citing *Cooperation Proclamation*, *supra note* 219).

277 *See Moore*, 2012 WL 607412, at *12 ("[T]he Court found it very helpful that the parties' ediscovery vendors were present and spoke at the court hearings where the ESI Protocol was discussed.").