

Mediated Investigative E-Discovery

Marian Riedy,¹ Suman Beros,² & Kim Sperduto³

ABSTRACT

The standard methodology used by parties today in formulating and responding to discovery requests is “keyword searching.” But this methodology is ineffective: the percentage of all discoverable information that is actually retrieved using keyword searches is demonstrably low. Simply broadening the search by using more general search terms may increase the retrieval percentage, but at an inordinate cost to both the requesting and producing party. When counsel crafts keyword searches the result may be ineffective in another sense: a successful deployment of the methodology may require the input of experts in the field. The solution to the technical problems with keyword searches may be for parties to propose alternative search and retrieval methodologies in discovery. But this would not solve a more fundamental problem with e-discovery today: increasingly contentious and expensive disputes over e-discovery, often centered over the search and retrieval methodology, the resolution of which is a substantial drain on judicial resources. The authors propose a novel process for conducting discovery and resolving disputes: mediated investigative e-discovery, in which a digital forensic investigator actually conducts the search for electronically stored information (ESI) at the direction of the parties, as well as facilitates agreements on production between the parties. Mediated investigative e-discovery holds the potential for efficiently achieving full and fair discovery.

1. Marian K. Riedy, Esq., has practiced as a litigator since graduating from Harvard Law School in 1981, has an MBA from Georgetown University, and teaches business law at Emporia State University.

2. Suman Beros, MBA, is an International Society of Forensic Computer Examiners Certified Computer Examiner, High Technology Crime Investigation Association member and an information technology executive experienced in startup and large corporate environments.

3. Kim Sperduto, Esq., is the managing partner of The Sperduto Law Firm, Washington, D.C.

I. INTRODUCTION

Since the 2006 amendments to the Federal Rules of Civil Procedure and the widespread adoption of keyword search techniques, electronic discovery has been cursed by the dual demons of unresponsive and over-responsive (data dump) production. Discovery disputes have escalated as a result. This article probes these issues and proposes a new paradigm for resolving e-discovery disputes in the real world crucible of high-stakes litigation.

The Blair and Maron study on the efficacy of keyword searching to retrieve documents relevant to a legal investigation is well known to those familiar with search and retrieval technologies in litigation.⁴ In the wake of a Bay Area Rapid Transit (BART) train accident, BART's defense counsel attempted to retrieve all documents relevant to their investigation using a keyword search. The database comprised approximately 40,000 documents, which represented roughly 350,000 pages of hard copy text.⁵ The search was conducted as an iterative process: the attorneys generated fifty-one different information requests, and each query was revised a number of times.⁶ By comparing the documents retrieved using keyword searching to all relevant documents identified through a manual search, the researchers concluded that the weighted average value of recall (percentage of all relevant documents retrieved) was approximately 20%.⁷

Manual review was an acceptable backup for the BART attorneys, given the number of documents to be searched. Printing and reading twenty-six million e-mail messages, however, would be inordinately costly and time-consuming.⁸ Leaving cost aside, manual review may not even be optimally productive in large-volume discovery cases. As has been noted, "there appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible—perhaps even

4. David C. Blair & M. E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 COMM. OF THE ACM 289 (1985). Though search algorithms and search strategies have surely improved since 1985, the Blair and Maron study is still referenced because "there has been little in the way of peer-reviewable research establishing the efficacy of various methods of automated content analysis, search, and retrieval as applied to a legal discovery context." The Sedona Conference® WG1, *The Sedona Conference® Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189, 215 (2007) [hereinafter *Best Practices Commentary*].

5. Blair & Maron, *supra* note 4.

6. *Id.* at 291.

7. *Id.* at 293.

8. See *United States v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 1 (D.D.C. 2006).

perfect—and constitutes the gold standard by which all searches should be measured.”⁹

Today, an internal review, as conducted by BART’s counsel, would be accomplished with the full panoply of tools used by a digital forensics investigator,¹⁰ increasing the recall percentage substantially over a keyword search, while confining manual review to a minimum, to control costs. A keyword search would probably be included, but it is merely the tip of the investigative iceberg. Thus, for counsel conducting an internal investigation today, the low retrieval percentage of the keyword search methodology illustrated in the Blair and Maron study can be largely ignored.¹¹

In making discovery requests and evaluating responses, litigators are generally relegated to just the tip of the iceberg. As discussed in Part II, the relative inefficacy of using keyword searches as the sole search and

9. *Best Practices Commentary*, supra note 4, at 199.

10. For example, a widely used AccessData’s Forensic ToolKit (FTK) organizes evidence items by File Status and File Category. The File Status classification includes Known File Filter (KFF) Alert Files, Bad Extension, Encrypted Files, From E-mail, Deleted Files, From Recycle Bin, Duplicate Items, OLE Subitems, Flagged Ignore, KFF Ignorable and Data Carved Files. By classifying all evidence items in this way the tool makes it possible for the investigator to prioritize the order of files to be examined and thus makes the examination more efficient. The bookmarking feature allows the investigator to easily revisit files that become relevant in light of subsequent evidence. By identifying duplicate files and known system and application files, the tool enables the investigator to take advantage of data reduction techniques and allows the investigator to skip over large number of files and to focus on user-generated files with potential evidentiary value. The File Category classification includes Documents, Spreadsheets, Databases, Graphics, Multimedia, E-mail Messages, Executables, Archives, Folders, Slack/Free Space, Other Known Type and Unknown Type enabling additional prioritization and efficiency. In addition to organizing evidence, FTK enables the investigator to search the evidence in a number of different ways, including by keywords. See AccessData-Forensic Toolkit 3.0, <http://www.accessdata.com/forensic toolkit.html>. Besides a general integrated software suite of tools exemplified here by FTK, a digital forensic investigator will typically use many specialized tools to support different forensic techniques, different types of investigations, operating and file system combinations, and many other considerations. A specialized tool could be used for specific forensic techniques (for example, data carver to uncover deleted data from the unallocated space or context triggered piecewise hashing to uncover even the smallest remnants of the known files). The tools could also be specialized to facilitate examination of specific categories of data (for example, email). Email tools facilitate analysis by creating visual representations of correspondent threads and clusters and allow visual navigation through such threads and clusters. Many excellent sources of information about commercial and open source forensic tools are available. For a concise and well-organized overview, see MICHAEL G. SOLOMON, DIANE BARRETT & NEIL BROOM, *COMPUTER FORENSICS JUMPSTART*, 161-192 (2005). Since digital forensics, like other information technology specializations, is a rapidly evolving area, for more up to date information consult online sources. See also Forensic Focus, <http://www.forensicfocus.com/computer-forensics-forums> (last visited Apr. 25, 2010) (displaying a general forensic forum and a good starting point to learn about the latest tools and techniques).

11. A full forensics evaluation would not necessarily be inexpensive, but the problem of the inefficacy of keyword searches alone for evaluating a complex “train wreck” of a lawsuit can be substantially overcome in an internal investigation.

retrieval methodology is a concern expressed in all quarters: by judges,¹² attorneys,¹³ commentators,¹⁴ and e-discovery consultants.¹⁵ Another significant problem with keyword searches—and one that will presumably increase in importance as litigators become more sophisticated in the use of the technology—is resolving disputes that arise as the parties propose alternative keywords. According to those studying these problems, in a nutshell, “there has to be a better way.”

In Part III, the “better ways” being proposed to solve the problems with the keyword search methodologies are critiqued. Described in Part IV is an alternative: mediated investigative e-discovery, which may be the most efficient means of attaining the objective of full and fair discovery.

II. KEYWORDS IN COURT

In one of her landmark *Zubulake* opinions, Judge Scheindlin observed that electronically stored information (“ESI”) “is frequently cheaper and easier to produce than paper evidence because it can be searched automatically,” such as by using “keyword” searches.¹⁶ Selecting keywords to identify requested information and to search for responsive information quickly became the status quo,¹⁷ in part because litigators were comfortable using the same methodology in discovery employed to find cases in Lexis or Westlaw or everything else with Google.

As keyword searches became the e-discovery tool of choice, so did the courts turn to this methodology as a means of resolving discovery issues. For example, courts direct the parties to meet and confer and agree on search terms,¹⁸ or, absent agreement, propose keywords for the court to determine which search(es) should be undertaken.¹⁹ Courts evaluate the

12. *E.g.*, *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008) (Facciola, J.).

13. *E.g.*, Mia Mazza, Emmalena K. Quesada, & Ashley L. Stenberg, *In Pursuit of FRCP 1: Creative Approaches to Cutting and Shifting Costs of Discovery of Electronically Stored Information*, 13 RICH. J.L. & TECH. 11 (2007), <http://law.richmond.edu/jolt/v13i3/article11.pdf>.

14. *See, e.g.*, *Best Practices Commentary*, *supra* note 4, at 215.

15. *E.g.*, Gregory L. Fordham, *Using Key Word Search Terms in E-Discovery and How They Relate to Issues of Responsiveness, Privilege, Evidentiary Standards and Rube Goldberg*, 15 RICH. J. L. & TECH. 8, 24-34 (2009), <http://law.richmond.edu/jolt/v15i3/article8.pdf>.

16. *Zubulake v. UBS Warburg, LLC*, 217 F.R.D. 309, 318 (S.D.N.Y. 2003) (citing Shira A. Scheindlin & Jeffrey Rabkin, *Electronic Discovery in Federal Civil Litigation: Is Rule 34 Up to the Task?*, 41 B.C. L. REV. 327, 335-341 (2000)).

17. *See, e.g.*, *Best Practices Commentary*, *supra* note 4, at 215.

18. *E.g.*, *Dunkin’ Donuts Franchised Rest. LLC v. Grand Cent. Donuts, Inc.*, No. CV 2007-4027, 2009 U.S. Dist. LEXIS 52261, at *14-15 (E.D.N.Y. June 19, 2009).

19. *E.g.*, *Bank of Mongolia v. M & P Global Fin. Servs., Inc.*, 258 F.R.D. 514 (S.D. Fla. 2009).

sufficiency of a party's response to discovery requests by reference to the particular keyword search conducted.²⁰ If a dispute arises as to whether the retrieved documents should be produced, the court may direct the responding party to conduct a keyword search for all documents which are responsive and relevant to the requests.²¹

Necessarily, then, courts are required to resolve search protocol issues when the parties cannot agree on keywords. Even a cursory reading of cases addressing e-discovery concerns illustrates that this task is neither useful nor appropriate for the judiciary to undertake. A representative case is *ClearOne Communications, Inc. v. Chiang*, a patent and copyright infringement action.²² Initially, pursuant to court order, the defendants' computers were imaged and the images placed in the custody of a third party.²³ Access to the data was not permitted until the parties had agreed upon a search protocol.²⁴ Agreement was reached on substantive search terms (names of specific individuals, technological references and terms related to the licensing of a particular code), but the parties could not agree on whether the terms should be connected disjunctively or conjunctively.²⁵ The court ruled as follows:

As to the "Name" terms, conjunctive search seems necessary. Otherwise, every occurrence of the "Name" terms will result in a positive hit, meaning that virtually every document in the electronic media will be identified as potentially responsive. In a relatively small business such as [defendant] WideBand, almost every document will refer to one of the key employees in the company. Requiring a hit of one "Name" term AND one "Tech" term will ensure that more responsive documents are flagged as potentially responsive.

However, as to the "License" terms, conjunctive search could be excessively narrow. Again, because WideBand is a relatively small company, licensing activity would be relatively small. By comparison, technology is the core of its business, so disjunctive use of the "Tech" terms would probably result in an excessive number of false positives.²⁶

20. *E.g.*, *Rexall Sundown, Inc. v. Perrigo Co. of S.C.*, No. CV 07-3397, 2008 U.S. Dist. LEXIS 47617, at *2 (E.D.N.Y. June 19, 2008); *Elliott v. U.S. Att'y Gen.*, No. 06-1128, 2006 U.S. Dist. LEXIS 80204, at *7-9 (D.D.C. Nov. 2, 2006).

21. *J. C. Assoc. v. Fid. & Guar. Ins. Co.*, No. 01-2437, 2006 U.S. Dist. LEXIS 32919, at *1-4 (D.D.C. 2006).

22. Amended Complaint at *10-16, *ClearOne Commc'ns, Inc. v. Chiang* No., 2:07 CV 37 TC, 2008 U.S. Dist. Ct. Pleadings LEXIS 1441 (D. Utah May 21, 2008).

23. *ClearOne Commc'ns, Inc. v. Chiang*, No. 2:07 CV 37 TC, 2008 U.S. Dist. LEXIS 27617, at *3 (D. Utah Apr. 1, 2008).

24. *Id.* at *3-4.

25. *Id.* at *4-5.

26. *Id.* at *5, 6.

Similarly, in *Asarco v. United States Environmental Protection Agency*, the plaintiff objected to EPA's use of only one term, "recontamination," to search its electronic files.²⁷ The court agreed and ordered that another search be conducted using the terms "recontaminate," "recontaminat," "recontamination," and "contaminate again." Given today's crowded dockets, is fashioning search terms really a function that our federal and state courts should be called on to perform?

Analyzing keyword searches is a strain on judicial resources, and may also be outside the ken of judicial abilities. Courts now evaluate keyword search proposals just as they would document requests: determining whether the proposed keywords are relevant to the claims or defenses or, if arguably relevant,²⁸ whether objectionable because the resulting retrieval would likely be overbroad.²⁹ Conversely, so long as the keywords chosen by the responding party to identify responsive documents are "reasonable," and the search "systematic" and "could be reasonably expected to produce the information requested," the keyword search passes muster.³⁰

Concern has been raised that the courts' experience in ruling on discovery issues regarding the content of paper documents does not directly translate to proposed keyword searches. In *United States v. O'Keefe*,³¹ in which the Government charged an employee of the Department of State for allegedly accepting "gifts" in exchange for expediting visa applications for employees of the defendant company, the defendants challenged the Government's choice of search terms used in responding to discovery requests. Magistrate Judge Facciola ruled that: "if defendants are going to contend that the search terms used by the government were insufficient, they will have to specifically so contend in a motion to compel and their contention must be based on evidence that meets the requirements of Rule

27. *Asarco v. EPA*, No. 08-1332 (FGS/JMF), 2009 U.S. Dist. LEXIS 37182, at *3 (D.D.C. Apr. 28, 2009). Though asked to rule on the search terms, Judge Facciola accompanied his order with a comment derogating the use of search terms at all, as discussed herein.

28. Of course, the responding party may also raise objections based on privilege and confidentiality. E-discovery in this arena raises a host of issues which this article does not address. See, e.g., *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008) (Grimm, J.); John M. Facciola & Jonathan M. Redgrave, *Asserting and Challenging Privilege Claims in Modern Litigation: The Facciola-Redgrave Framework*, 4 FED. CTS. L. REV. 19 (2010).

29. See, e.g., *Capitol Records, Inc. v. MP3Tunes, LLC*, 261 F.R.D. 44, 50, 53 (S.D.N.Y. 2009) ("Each of the thirty disputed terms is relevant to one or more of these additional claims." The court later said, "[T]he search terms that MP3tunes proposes clearly are overbroad . . .").

30. See *Anderson v. U.S. Dep't of State*, No. 09-569 (ESH), 2009 U.S. Dist. LEXIS 95634, at *13-14 (D.D.C. Oct. 14, 2009).

31. 537 F. Supp. 2d 14, 23-24 (D.D.C. 2008).

702 of the Federal Rules of Evidence.”³² In reaching this conclusion, Judge Facciola reasoned:

Whether search terms or “keywords” will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics. Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread. 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.³³

This language suggests that expert witnesses may become an inevitable and necessary ingredient in resolving complex e-discovery disputes.

In a subsequent decision, Judge Facciola observed that a party may also be required to corroborate—with appropriate technical evidence—the validity of the search and retrieval technology chosen, whether it be a keyword search or an alternative technology.³⁴ In *Victor Stanley, Inc. v. Creative Pipe, Inc.*, Judge Paul W. Grimm echoed Judge Facciola’s observations in assessing the “reasonableness” of the efforts by a party to prevent the inadvertent disclosure of privileged information.³⁵ Courts in other jurisdictions have also noted the need for expert evidence in resolving discovery disputes based on electronic search and retrieval issues,³⁶ or simply ordered that search protocols be developed by the parties’ respective experts.³⁷

32. *Id.* at *24.

33. *Id.*

34. *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008) (requiring Equity Analytics to submit an affidavit from its examiner describing in detail how the search would be conducted and explaining why the limitations proposed would be unlikely to capture all the information sought).

35. *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 262 (D. Md. 2008) (“[d]efendants neither identified the keywords selected nor the qualifications of the persons who selected them to design a proper search; they failed to demonstrate that there was quality-assurance testing; and when their production was challenged by the Plaintiff, they failed to carry their burden of explaining what they had done and why it was sufficient.”).

36. *William A. Gross Constr. Assoc., Inc. v. Am. Mfr. Mut. Ins. Co.*, 256 F.R.D. 134, 135 (S.D.N.Y. 2009). Related to the issue of validating search and retrieval methodologies by expert evidence is showing that a proposed sampling of a database or the custodian to determine whether a further search is indicated, meets statistically-valid criteria. *See, e.g., In re Vioxx Prod. Liab. Litig. Steering Comm. v. Merck*, No. 06-30378, No. 06-30379, 2006 U.S. App. LEXIS 27587, at *6-7 (5th Cir. May 25, 2006); *Am. Nat’l Bank & Trust Co. v. Equitable Life Assurance Soc’y of the U.S.*, 406 F.3d 867, 879 (7th Cir. 2005).

37. *John B. v. Goetz*, 531 F.3d 448, 453 (6th Cir. 2008) (explaining the district court held an “experts only” conference where it directed the parties’ computer experts to confer to develop a protocol to address problems with electronic discovery).

The clear implication from these decisions is that attorneys would be well-advised to seek expert advice in crafting and responding to discovery requests, not only to validate a search and retrieval methodology should a dispute arise, but also because counsel, like the courts, do not have the expertise in many circumstances to cope with e-discovery. The technology is simply “beyond the ken” of a layman. In particular, litigants are being encouraged to re-examine the prevailing use of keyword searches because of the risk of low retrieval, as demonstrated in the Blair and Maron study,³⁸ and/or the recovery of massive amounts of irrelevant data.³⁹ Indeed, in some circles “keyword searches are no longer the favored methodology,”⁴⁰ because “concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results.”⁴¹ In other words, there is a stark difference between the e-discovery keyword orthodoxy today, and the apparent tenets and rituals of the emerging new creed.

In lieu of producing an expert, or in combination therewith, the prevailing advice—or directive—to litigants is to cooperate and agree on search and retrieval methodologies. Hence:

Electronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI. Moreover, where counsel are using keyword searches for retrieval of ESI, they at a minimum must carefully craft the appropriate keywords, with input from the ESI’s custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of “false positives.” It is time that the Bar—even those lawyers who did not come of age in the computer era—understand this.⁴²

The Sedona Conference has issued a “Cooperation Proclamation,” which strongly encourages litigators to cooperate on e-discovery issues in

38. Blair & Maron, *supra* note 4.

39. “It is more likely than not that the search terms may produce documents that will lack any relevance to plaintiffs’ claims since no one can pretend that the search terms are such finely honed instruments that they will only produce what is relevant.” *In re Rail Freight Fuel Surcharge Antitrust Litig.*, Misc. No. 07-489 (PLF/JMF/AK), 2009 U.S. Dist. LEXIS 99187 at *32 (D.D.C. Oct. 23, 2009).

40. *Asarco v. EPA*, No. 08-1332 (FGS/JMF), 2009 U.S. Dist. LEXIS 37182, at *7 (D.D.C. Apr. 28, 2009).

41. *Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, 242 F.R.D. 139, 148 (D.D.C. 2007) (citation omitted).

42. *William A. Gross*, 256 F.R.D. at 136. *See also* *Novelty, Inc. v. Mtn. View Mktg.*, 1:07-cv-01229-SEB-JMS, 2009 U.S. Dist. LEXIS 98592, at *16, n.9 (“[P]arties can, and should, work together to develop keywords when large amounts of electronic data must be searched for the sake of efficiency and cost-savings.”).

order to reduce the “serious burden to the American judicial system” associated with the costs of e-discovery in an adversarial system.⁴³ The Proclamation argues that “economy and logic” compel a “paradigm shift” in discovery, from an adversarial process to one of cooperation to “promote open and forthright information sharing, dialogue (internal and external), training, and the development of practical tools to facilitate cooperative, collaborative, transparent discovery.”⁴⁴

III. CONFLICTING OBJECTIVES

Litigators are being charged with the tasks of validating proposed search methodologies with expert evidence, utilizing specialized expertise and emerging technologies to increase the efficiency and effectiveness of retrieval,⁴⁵ and cooperating more during discovery. Such marching orders, to some extent, are discordant.

As Judge Grimm observed, “It cannot credibly be denied that resolving contested issues of whether a particular search and information retrieval method was appropriate—in the context of a motion to compel or motion for protective order—involves scientific, technical or specialized information.”⁴⁶ Nor can it credibly be denied that increasing the use of expert testimony to validate search requests will increase costs. It is simply a fact that expert witnesses cost money. And given the multidisciplinary nature of electronic search and retrieval, involving “the sciences of computer technology, statistics and linguistics,” it can reasonably be assumed that no single expert will suffice to validate a particular search and retrieval request. For example, though a technology expert could opine as to the appropriate technology to be employed, a linguist may be required to explain why “concepts,” not apparently relevant to any issue, would be reasonably calculated to lead to the discovery of admissible evidence. Furthermore, if the judiciary fully retreats from assessing proposed searches, all electronic discovery requests would have to be validated by

43. THE SEDONA CONFERENCE, THE SEDONA CONFERENCE® COOPERATION PROCLAMATION 1 (2008), available at http://www.thesedonaconference.org/content/tsc_cooperation_proclamation/proclamation.pdf.

44. *Id.* at 1. The Proclamation also posits that the Federal Rules of Civil Procedure “mandate” that the parties act cooperatively during discovery.

45. “[L]awyers must embrace creative, technological approaches to grappling with knowledge management as information inflation continues apace. Failure to do so will severely hamper the legal profession’s ability to meaningfully retrieve and process evidence.” Gary L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J. L. & TECH. 10, at 7 (2007), <http://jolt.richmond.edu/v13i3/article10.pdf>.

46. *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 261 n.9 (D. Md. 2008).

expert evidence, unless the parties agreed on the topics, method and manner of the search.

Adding experts to the process also reduces the probability that the parties will be able to “cooperate” and agree on a search methodology, *unless the respective experts agree*. It would be rare for a party to engage and work with experts to formulate a search methodology, then concede to an alternative proposed by the opposing party’s expert.

The interplay between the objectives of validating search requests and increasing the efficiency and effectiveness of retrieval is also problematic. Though it may be assumed that the use of experts will improve a *proposed* search methodology, problems may arise in terms of execution. Even experts must rely on an iterative process effectively to “filter” discoverable from the available universe of information.⁴⁷ Traditionally, the discovery process just does not work this way: each party serves a comprehensive set of requests, reviews the responses and objections, serves supplemental requests, and perhaps serves a second set of requests in the event an unanticipated issue arises. Even a successful Google search for one specific item of information is usually more iterative than discovery. Certainly the discovery process today is coarse and simplistic compared to the successive sweeps and iterations a digital investigator would perform if given full access to ESI.

Cooperation, then, becomes the key to avoid an expensive “battle of the experts” and to improve the efficiency and efficacy of e-discovery. The deep-seated and systemic adversarial character of litigation and the “duty” of zealous representation, however, are major barriers to achieving this objective. Critics condemn litigators for “hiding the ball.”⁴⁸ Clearly, it is unethical to hide the ball if the other party has requested it, but it is another thing entirely to volunteer to the opposing party that your client has a ball. Otherwise stated, it is perfectly appropriate and necessary to identify all the client’s data custodians, just as one would identify all known fact witnesses. Counsel, however, is not expected to suggest what information the opposing party should look for—proposed search terms, for example—just as counsel is not required to suggest to the opposition what questions should be asked of a witness during deposition.⁴⁹

47. See THE SEDONA CONFERENCE, COMMENTARY ON ACHIEVING QUALITY IN THE E-DISCOVERY PROCESS 15 (2009), available at http://thesedonaconference.org/dltForm?did=Achieving_Quality.pdf (“The filtering process should be iterative and needs to be repeated until the desired goals are met.”).

48. THE SEDONA CONFERENCE COOPERATION PROCLAMATION, *supra* note 43, at 1.

49. Indeed, it is curious that the courts regularly require a party to identify what search terms were used to identify responsive documents, for unless the party is “hiding the ball,” those terms can go far towards reflecting counsel’s thought processes. In *Smith v. Life Investors Ins.*

A party could cooperate by simply agreeing to a search methodology proposed by the opposing party; such cooperation could be costly and time-consuming. Further, parties often have legitimate disputes about whether requested information is discoverable. Simply postponing that dispute until after the opposing party has retrieved and reviewed information—and allowing full “cooperative” discovery—is not an option. For example: suppose a plaintiff seeks discovery held by a third party alleged by the plaintiff to be the defendant’s agent. The defendant can object on the ground that the third party is an independent contractor. Under this scenario, a defendant has no legal obligation to obtain documents from the independent contractor because the requested information is not within the defendant’s “possession, custody, or control.” Nonetheless, a defendant cooperates and allows the discovery. Indisputably, relevant evidence adverse to the defendant on an issue unrelated to the agency/contractor distinction emerges. The client is decidedly unhappy.

Simply conceding to the opposing party’s requested search methodology does not guarantee an efficient retrieval. For example, in *CBT Flint Partners, LLC v. Return Path, Inc. et al*, one of the defendants performed a keyword search for responsive information and produced 1.4 million documents.⁵⁰ The plaintiff’s complaint that this production constituted a “document dump” was rejected because it chose the search terms used by the defendant.⁵¹

In addition to the substantive barriers to the proposed cooperative discovery are the practical issues. For example, the Sedona Cooperation Proclamation recommends that the parties name ESI discovery “point persons” to assist counsel in identifying discovery requests, and that they cooperate in “jointly developing automated search and retrieval methodologies.”⁵² Identified “point persons” could presumably allow the parties to craft requests and evaluate responses more rapidly, leading to a more iterative and effective search and retrieval process. Any issue of real substance, however, would still have to be resolved by senior counsel, with the assistance of junior counsel most familiar with the case file, and IT consultants. As noted by courts and commentators, attorneys are no more adept with these issues than are judges. How is counsel expected to

Co. of Am., 2009 U.S. Dist. LEXIS 58261, at *20 (W.D. Pa. 2009), the court rejected the argument that search terms unilaterally chosen by counsel constituted attorney work product, but the argument seems to have some merit.

50. No. 1:07-CV-1822-TWT, 2008 U.S. Dist. LEXIS 84189, at *5 (N.D. Ga. Aug. 7, 2008).

51. *Id.* at *7-8.

52. *The Sedona Conference Cooperation Proclamation*, *supra* note 43, at 2.

cooperate and agree on “concepts” or “fuzzy logic” when the task is difficult enough using search terms, which are at least loosely tethered to a “relevant” word or phrase and a more familiar tool?⁵³ The answer is for counsel to turn to experts, an exercise that seems unlikely to stimulate cooperation.

Another suggestion in the Sedona Cooperation Proclamation is that the parties consider court-appointed experts, volunteer mediators, or formal ADR programs to resolve discovery disputes.⁵⁴ Mediation may be effective in resolving discovery disputes, like it has been in settling cases. But in the “standard” mediation, the mediator facilitates dispute resolution, while the parties conduct their own search and retrieval.⁵⁵ To the extent that parties are “hiding the ball,” the standard mediation process does not solve the problem, because the mediator is only aware of the information the parties voluntarily disclose, and has no leverage to induce information exchange. Further, assuming that each party has already undertaken an internal investigation, whatever search and retrieval methodology is agreed to and implemented in the mediation process may be redundant and an unnecessary expense.

In the alternative approach described below, the mediator actually conducts the search, as directed by counsel, as well as facilitates agreements on production between the parties. This approach—mediated investigative e-discovery—holds the potential for efficiently achieving full and fair discovery.

53. It should also be noted that, for the present, effective concept search and retrieval software is expensive.

54. *The Sedona Conference Cooperation Proclamation*, *supra* note 43, at 2.

55. See, e.g., Allison Skinner, *How to Prepare an E-Mediation Statement for Resolving E-Discovery Disputes*, SIROTE (Sirote & Permutt, PC, Birmingham, Ala.), <http://smu-ecommerce.gardere.com/allison%20skinner%20preparing%20for%20e-mediation%20discovery.pdf> (last visited March 28, 2010).



IV. MEDIATED INVESTIGATIVE E-DISCOVERY

The animating principle of mediated investigative e-discovery is that discovery can be more efficient and effective, and completed without the need for motions to compel and intervention by the court, when it is assisted by a neutral third party employing the skills of both a trained digital investigator and a mediator. In the investigative role, the mediator-investigator can effectively search and retrieve relevant information, similar to an in-house expert, but at less cost, as the process is more effective and both parties share the expense. By direct involvement in the search, the mediator-investigator becomes knowledgeable about the strengths and weaknesses of both parties' position in the discovery dispute, and is thereby armed with information to assist in negotiating a solution.

Not every case needs a trained investigator to search for discoverable information. If the key allegation is that the boss sent a series of sexually-explicit e-mail messages to a secretary from a company-issued Blackberry, the "for what and where" to search is obvious. But if the case involves the proverbial "train wreck," the cause of which is unknown, the party preparing e-discovery requests faces a daunting task.

Not every discovery dispute can be resolved with the assistance of a trained mediator. Every judge, magistrate, and practitioner knows, however, that mediation can be extremely effective. For example, when counsel are close to what they believe is a fair agreement, but one client is taking an unreasonably inflexible position, a neutral third party can often convince the client that settlement is the better alternative to a trial.

Assuming, then, a case for which both parties anticipate significant and complex e-discovery, the process of mediated investigative e-discovery

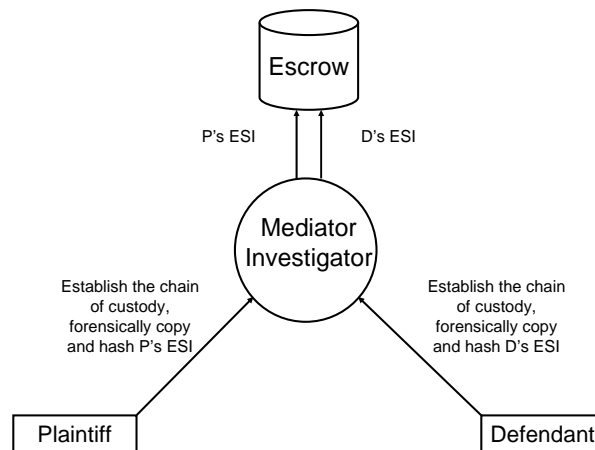
56. Tom Fishburne, *Opposing Counsel Counseling*, CASE IN POINT, Feb. 12, 2009, <http://www.casecentral.com/case-in-point/?m=200910> (last visited March 28, 2010).

and its advantages are described in Sections A and B below. Potential criticisms to the use of this process, and responses to those objections, are discussed in Section C.

A. Mediated Investigative E-Discovery

Mediated investigative e-discovery includes three necessary components: (1) protecting and preserving the data, (2) conducting the investigation, and (3) mediating any disputes about production of the retrieved information.

Figure 1: Protecting and Preserving the Data



In the typical case, the mediator-investigator would first obtain ESI from the custodians on which the parties have agreed during the “meet and confer” required by Fed. R. Civ. P. 26(f),⁵⁷ establish the chain of custody, forensically copy and hash the ESI⁵⁸ and place it in escrow. The preferred

57. Alternatively, if the parties cannot agree on the scope of ESI to be preserved or the custodians from which data will be obtained, the mediator/investigator can assist in resolving that dispute. Either way, to obtain the maximum benefit from the process, the mediator/investigator should be engaged very early in the discovery process.

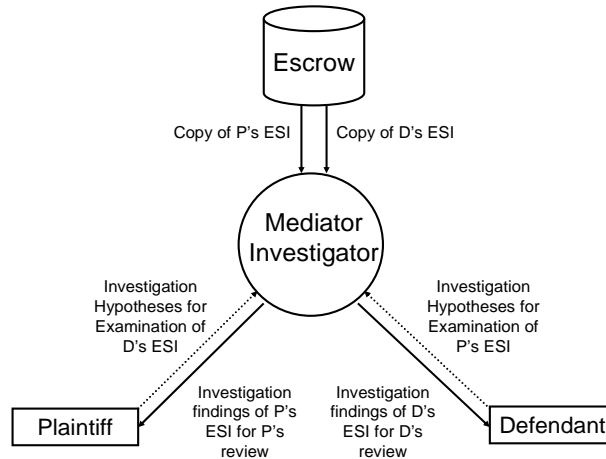
58. Hashing is the process of using a mathematical algorithm against data to produce a numeric value that is representative of that data. Even the smallest change in the data, for example a single bit change, results in a changed hash. Computing a hash of a data file and comparing it to a previously computed hash of that data file is a widely accepted method to establish or confirm that the subject data file is unchanged. If the hashes match, the data is the same. If the hashes do not match, the data is not the same. Because of this characteristic, hashes are frequently referred to as “digital fingerprints.”

methodology is to forensically copy⁵⁹ all ESI from all custodians, although in some situations capturing only the normal active files may suffice.⁶⁰ If only the normal active files are to be captured, care should be exercised to also secure those files' metadata and to protect the target ESI from any changes, inadvertent or otherwise, during the capture process. Regardless of whether only the normal active files are captured or whether the media is forensically copied, all obtained ESI should be hashed and escrowed. In the mediation agreement, the parties can define the specific circumstances for release of ESI from the escrow. If the parties cannot agree on release terms, the default should be that the final arbiter is the court.

59. A forensic copy of an electronic data storage medium is an identical bit by bit copy of that medium, containing the complete contents and structure of that data storage medium or device, such as a hard drive, USB flash drive, compact disc (CD), or digital versatile disc (DVD), based on the physical sectors on the medium and ignoring the file system. Because a forensic copy is an identical copy of a data storage medium, it includes any normal files, any temporary files, any deleted files, and any file fragments whether found in files, in file slack or in the unallocated space on the original medium. A forensic copy can be stored on the medium of the same type that is at least as big as the original medium. This type of forensic copy is also known as a forensic clone. More likely, a forensic copy will be stored in the form of an image file (a forensic image) that can be stored on different media types for archiving, restored on different media types for examination or directly accessed for examination using special applications, such as AccessData's Forensic Toolkit (FTK).

60. See *Craig Ball on e-Discovery and Forensics Trends in 2009* (Fios, Inc. webcast on Feb. 24, 2009), <http://www.fiosinc.com/e-discovery-knowledge-center/electronic-discovery-webcast.aspx?id=491> (explaining that (1) the forensic acquisition is appropriate when spoliation or fraud are suspected; (2) there is evidence of system intrusion; (3) there is a close nexus between the system or the device and the cause of action; (4) when the status or conduct of the ESI custodian is critical; and, (5) as inexpensive insurance, if C-level executives are implicated).

Figure 2: Conducting the Investigation



After consulting with a plaintiff⁶¹ to understand the substantive issues and the plaintiff's theories of the case, the mediator-investigator performs an investigation⁶² of the defendant's ESI⁶³ based on the plaintiff's claims/defenses. The plaintiff may choose to have the mediator-investigator search for ESI relevant to all issues and theories of the case,⁶⁴ or first target the strongest theory, based on the plaintiff's initial assessment. For example, in the BART case referenced in Part I, the target of the first search might be maintenance, based on counsel's experience that maintenance errors are a likely cause of the wreck.

61. The mediator-investigator does not need to and should not have any contact with the ultimate clients, either plaintiff or defendant, without their respective counsel. The exception is incidental contact during data collection. In any event, the mediator-investigator should never discuss the case with the parties outside the presence of their counsel.

62. "A digital investigation is a process where we develop and test hypotheses that answer questions about digital events. This is done using the scientific method where we develop a hypothesis using evidence that we find and then test the hypothesis by looking for additional evidence that shows the hypothesis is impossible. Digital evidence is a digital object that contains reliable information that supports or refutes a hypothesis." BRIAN CARRIER, *FILE SYSTEM FORENSIC ANALYSIS* 4 (2005).

63. The mediator-investigator works with a copy of the ESI in escrow.

64. See MARIAN K. RIEDY, SUMAN BEROS, & KIM SPERDUTO, *LITIGATING WITH ELECTRONICALLY STORED INFORMATION* 91-107 (2007), for the four approaches to identifying and requesting the relevant ESI. These approaches are the "Where" or the Computing Environment Model, the Data "Checklist" Model, the "Life cycle" Model and the Revised "refer or relate" Model.

To further control cost, the parties may initially agree to limit the mediator's investigation to a budgeted number of hours.⁶⁵

Plaintiff, working with the mediator-investigator, will develop a set of "discovery hypotheses" to be tested during the investigation. The "discovery hypotheses" related to maintenance could be, for example, that relevant ESI might be found in maintenance logs, parts acquisition/withdrawal records, and any correspondence about maintenance for the relevant period. The mediator-investigator would test these discovery hypotheses by searching for all ESI that supports or refutes Plaintiff's theory that the maintenance was faulty.

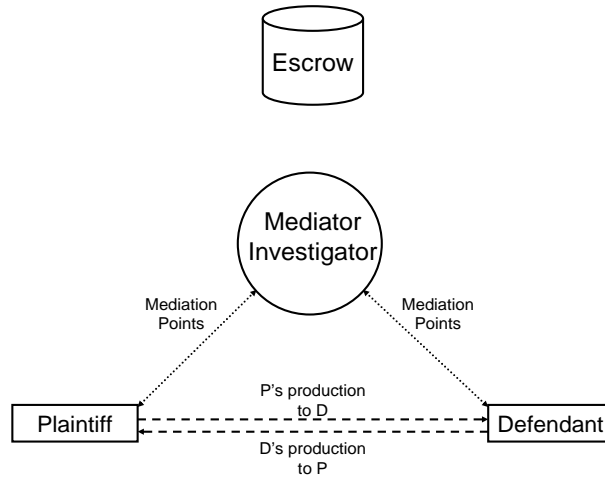
The mediator-investigator then performs an initial sweep of the data using standard data reduction techniques (eliminating applications, indexing data types, etc.) and tests the discovery hypotheses. The mediator-investigator is free to use any appropriate methodologies and techniques during the investigation. The process is iterative, and continues until the mediator-investigator is confident that the reliable information⁶⁶, which supports or refutes the theory, has been found. Ultimately, counsel decides, following production and based on an analysis of the information retrieved, whether the search has been sufficiently exhaustive.

The mediator-investigator then provides the retrieved information to Defendant, who can agree to produce or withhold it based on the standard objections, including relevance, privilege, and attorney work-product. The investigative process is replicated on Plaintiff's data on behalf of Defendant.

65. Applying the BART example to a budgeted hours arrangement, the mediator would probably triage the examination to focus first on the maintenance logs, and then prioritize examination of the parts acquisition/withdrawal records and any correspondence about maintenance, as well as, assess the reliability of such records if the budgeted time allows.

66. Digital data is typically stored in files, which are stored in file systems. In order to store and retrieve such data files effectively and efficiently, a number of data structures, in addition to the file content, are created and stored in a file system: file name, location of the file content and temporal data, for example. Whereas file name and the location of the file content are essential for a functioning file system, the temporal data is not. Because the temporal data is non-essential, it is less reliable than the essential data. Non-essential data can be changed intentionally or inadvertently without impacting the functioning of the file system. Digital forensics professionals are trained to evaluate data's reliability and look for corroborating evidence when confronted with non-essential or less reliable data.

Figure 3: Mediating Disputes about Producing the Retrieved Information



Once information is retrieved, negotiations between the parties, with the mediator-investigator operating in the more traditional role of mediator, may occur. The mediator knows the strengths and weaknesses of both parties' arguments and can assist them in exchanging bargaining chips—retrieved information—as a method of resolving disputes.

For most disputes, a digital forensic investigator trained in mediation could successfully fulfill the role of mediator-investigator. In other cases, where due to the scope of the ESI to be investigated a single professional may be inadequate, a team approach led by either a digital forensic investigator or an experienced mediator may be more appropriate.⁶⁷

B. Advantages of Mediated Investigative E-Discovery

1. Minimizes Preservation Issues

When the mediator-investigator is engaged early in the discovery process, any issues arising from the destruction, loss, or alteration of relevant or otherwise discoverable information are foreclosed. Because the mediator-investigator establishes the chain of custody, forensically copies,

67. Larger mediator-investigator teams may find it advantageous to organize themselves into two teams, each team focused on one of the parties' ESI, and both teams reporting to a senior mediator-investigator overseeing the process. Such organization would mitigate any perceived conflicts of interest, as well as facilitate quality control by having the teams validate each other's findings.

hashes, and retains the parties' ESI in escrow *as agreed by the parties*, the process effectively precludes any subsequent charges of inadvertent or intentional spoliation, and allows for third party quality control if the mediator-investigator's findings are later challenged. Both parties are also protected from the inadvertent loss of any of *their own* data that could be significant to a claim or defense. Finally, the process ensures that the basis for authenticating ESI has been established should the case be tried.⁶⁸

2. Improves the Effectiveness of Discovery

Every practitioner has nightmares about failing to obtain some key piece of evidence from the opposing party. Given the volume of ESI, and the present reliance on keyword searches, a relatively ineffective search technique, those nightmares should be recurring. Engaging a trained digital investigator to search the opponent's ESI would reduce such legitimate concerns.

3. Reduces Costs

Mediated investigative e-discovery allows and encourages the parties to conclude discovery sooner rather than later by removing some of the ways to "hide the ball" under the guise of technology, and by focusing on finding the "needle in the haystack." Mediated investigative e-discovery enables the parties to search for evidence of the "needle," rather than concentrating on the mechanics and the review and production of the universe of responsive ESI. Because the goal of this e-discovery approach is not directed to the production of the "haystack" of responsive ESI, a significant amount of expensive review time and effort can be saved by both parties.

Because mediated investigative e-discovery is also focused on identifying and triaging specific evidence that either supports or refutes a specific claim or defense, the process is more effective and efficient in several ways. Instead of reviewing the universe of responsive ESI for production to the other party, the responding party can focus the review on the specific subset of the ESI that the mediator-investigator has helped to identify and that directly supports or refutes the other party's claim. Second, the requesting party does not have to review the "haystack" of ESI to search for the "needle." The mediator-investigator has already helped to accomplish this by using the most appropriate tools and methodology, while

68. Appropriately preserving ESI and documenting the chain of custody is the necessary first step in establishing authenticity. *See RIEDY ET AL.*, *supra* note 64, at 187-195.

concurrently protecting the producing party's right to object. As previously discussed, the current e-discovery approach allows only a grossly iterative process resulting in inefficient use of data. In contrast, mediated investigative e-discovery, because of the mediator-investigator's unfettered access to ESI, permits efficiencies stemming from continuity and iteration.

The mediated investigative approach also lends itself to triage and thus the most economical use of limited resources. As the cost of data storage continues to decline and the typical keyword search increasingly returns irrelevant data dumps, the need for triage becomes more acute. The mediator may be asked to proceed with the investigation beyond triage, or not, since the initial triage, which safely captures and protects the "truth" in the escrowed ESI, may be sufficient.

Finally, mediated investigative e-discovery can be further managed through iterative and incremental cycles by setting limits on the amount of time and effort that the mediator-investigator will expend assisting counsel in pursuing specific hypotheses.

C. Criticisms and Responses

The main criticism to the mediated investigative e-discovery approach would be the loss of control: litigants allow a third party to conduct their discovery, and give that third party unfettered access to their own data.

The first concern can be readily eliminated, as it is simply a matter of perception. First, the mediator-investigator is at all times conducting the search *as directed by counsel*. Based on his or her expertise, the mediator-investigator will make informed recommendations about searching for ESI, but the final decision always rests with counsel. Second, counsel has much more access to—and therefore more control over—the opposing party's ESI by having a trained investigator review that data than counsel would have by using the timeworn techniques of interrogatories and requests for production.

The second criticism is more significant. More so than in a standard mediation, the entire process is heavily dependent on the skills, and in particular, the trustworthiness of the mediator-investigator. Typically, communications between a party and the mediator are privileged and protected from discovery by the opposing party or in any subsequent proceeding.⁶⁹ Information which the mediator-investigator has retrieved from a party's data would not qualify for such protection. The parties could (and no doubt would) require the mediator-investigator to enter into a confidentiality agreement providing that all retrieved information which a

69. See, e.g., MD. R. 17-102(e) & 17-109.

party does not agree to produce remains confidential, and further prevents the mediator-investigator from being compelled to disclose any such information in any proceeding. Depending on the applicable rules, such an agreement could be enforceable against third parties. For example, the Maryland rule provides:

The written agreement [executed by the neutral] may include provisions stating that the expert may not disclose or be compelled to disclose any communications related to the alternative dispute resolution proceeding in any judicial, administrative, or other proceedings. Communications related to the alternative dispute resolution proceeding that are confidential under an agreement allowed by this subsection are privileged and not subject to discovery, but information otherwise admissible or subject to discovery does not become inadmissible or protected from disclosure solely by reason of its use related to the alternative dispute resolution proceeding.⁷⁰

Arguably, the retrieved information would be “related to the alternative dispute resolution proceeding” given the design of mediated investigative e-discovery. But this argument has not yet been tested. Assuming that information retrieved by the mediator-investigator would not be protected from disclosure by statute, the risk of disclosure is not necessarily greater than is posed by any person or entity involved in the discovery process. Given sufficient inducement, a vendor engaged by a party, or the party’s paralegal, might be as likely to disclose information to the opposing party as is the mediator-investigator. At least the parties can be assured that information will not “disappear” if a mediator-investigator is involved, because all information adduced in discovery can be compared to the forensically copied ESI in escrow.

Another criticism may arise from the possibility that the mediator-investigator would identify evidence of crime during the investigation of ESI. If mere possession of that evidence is a crime (e.g., child pornography), the mediator-investigator is required to report such a finding to the proper authorities. This inherent risk, however, exists when utilizing a traditional e-discovery specialist or even internal IT staff. All bear the same responsibility to advise law enforcement of criminal possession.

A final objection may be counsel’s uneasiness with relying so heavily on an IT professional. Get over it. With today’s information universe so heavily dominated by ESI, the perfectly wrong response is to refuse the assistance of those who are conversant with the technologies of ESI.

A man is flying in a hot air balloon and realizes that he is lost. He reduces his altitude and spots a man down below. He lowers the balloon further and says, “Excuse me, can you tell me where I am?”

70. MD. R. 17-105.1(c)(2).

The man below says, “Yes. You are in a hot air balloon, hovering 30 feet above this field.

“You must work in Information Technology,” says the balloonist.

“I do,” replies the man, “How did you know?”

“Well,” says the balloonist, “everything you have told me is technically correct, but it’s no use to anyone.”

The man below says, “You must be a lawyer.”

“I am,” replies the balloonist, “but how did you know?”

“Well,” says the man, “You don’t know where you are, or where you’re going, but you expect me to be able to help. You’re in the same position you were before we met, but now, it’s my fault.”⁷¹

V. CONCLUSION

The initial promise of e-discovery—that it would be quicker, easier, and less expensive than discovery of paper documents—has clearly not been met. The discovery process, itself, must change to reduce cost and increase the efficiency of e-discovery, and rescue courts from a deluge of discovery disputes. Courts’ time is better expended on other matters. As Judge Scheindlin observed, after noting that she and her clerks spent three hundred hours resolving a discovery dispute: “My point is only that sanctions motions, and the behavior that caused them to be made, divert court time from other important duties—namely deciding cases on the merits.”⁷² Changing the nature of the American legal system from an adversarial to a “cooperative” one might just be too much to ask or to expect. Other solutions may be worth considering, and given its potential advantages, mediated investigative e-discovery should be high on the list.

71. This anecdote has been floating around on the Internet for many years without any claim to ownership.

72. *Pension Comm. of the Univ. of Montreal Pension Plan v. Banc of Am. Sec., LLC*, No. 05 Civ. 9016 (SAS), 2010 U.S. Dist. LEXIS 4546, at *33 n.56 (S.D.N.Y. Jan. 15, 2010).