The Sedona Conference Journal

Volume 11 2010

A Nutshell on Negotiating E-Discovery Search Protocols

Jason R. Baron & Edward C. Wolfe

Recommended Citation: Jason R. Baron & Edward C. Wolfe, *A Nutshell on Negotiating E-Discovery Search Protocols*, 11 SEDONA CONF. J. 229 (2010).

For this and additional publications see:
https://thesedonaconference.org/publications

A NUTSHELL ON NEGOTIATING E-DISCOVERY SEARCH PROTOCOLS

Jason R. Baron & *Edward C. Wolfe*[1]
Washington, DC *Bloomfield, MI*

The aim of this Nutshell is to provide practical guidance to counsel on the subject of conducting search negotiations as part of the Fed. R. Civ. P. 26(f) meet and confer process or in the context of negotiating a Rule 16 pre-trial order. The outline is intended to highlight key concepts and approaches that have emerged since the revisions to the Federal Rules of Civil Procedure went into effect in December 2006. Due to the pace of technological change and new case developments, the authors understand that what is said here may be in need of revision and updating shortly after its publication date, and thus counsel is cautioned to remain vigilant in keeping up with emerging case law and commentary in this area.

A. General Guidance

1. Conducting a Reasonable, Comprehensive Search Pursuant to Fed. R. Civ. P. 34

The Sedona Conference® *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery* ("*Sedona Search Commentary*") addressed the subject of how lawyers search for electronically stored information (ESI), providing a roadmap to counsel on the practical limitations of keyword searching as practiced by the profession, with practice points given on how to approach the task of satisfying one's professional obligations to conduct a reasonable, comprehensive search in response to a Rule 34 request for documents and ESI.[2] The *Sedona Search Commentary's* eight practice points, as set out and further explained therein, are as follows:[3]

> *Practice Point 1. In many settings involving electronically stored information, reliance solely on a manual search process for the purpose of finding responsive documents may be infeasible or unwarranted. In such cases, the use of automated search methods should be viewed as reasonable, valuable, and even necessary.*
>
> *Practice Point 2. Success in using any automated search method or technology will be enhanced by a well-thought out process with substantial human input on the front end.*

---

1 Jason R. Baron serves as Director of Litigation at the National Archives and Records Administration. Edward C. Wolfe was with the Office of the General Counsel of General Motors Company in Detroit for over 30 years. Both authors presently serve on the Steering Committee for The Sedona Conference® Working Group on Electronic Document Retention and Production (WG1). The views expressed herein are the authors alone and do not necessarily represent the views of the institutions with which they are or have been affiliated.

2 The Sedona Conference,® *The Sedona Conference® Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189 (2007), *available at* http://www.thesedonaconference.org/content/miscFiles/publications_html.

3 *Id.*

Practice Point 3. The choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed.

Practice Point 4. Parties should perform due diligence in choosing a particular information retrieval product or service from a vendor.

Practice Point 5. The use of search and information retrieval tools does not guarantee that all responsive documents will be identified in large data collections, due to characteristics of human language. Moreover, differing search methods may produce differing results, subject to a measure of statistical variation inherent in the science of information retrieval.

Practice Point 6. Parties should make a good faith attempt to collaborate on the use of particular search and information retrieval methods, tools, and protocols (including as to keywords, concepts, and other types of search parameters).

Practice Point 7. Parties should expect that their choice of search methodology will need to be explained, either formally or informally, in subsequent legal contexts (including in depositions, evidentiary proceedings, and trials).

Practice Point 8. Parties and the courts should be alert to new and evolving search and information retrieval methods.[4]

2. Employing Quality Control Techniques

Additional guidance on the subject of employing quality control techniques as part of the search process, including sampling and iterative methods, was provided in The Sedona Conference® *Commentary on Achieving Quality in the E-Discovery Process* ("*Sedona Achieving Quality Commentary*").[5] In particular, Principle 3 in the *Sedona Achieving Quality Commentary* states in relevant part that "[i]mplementing a well thought out e-discovery 'process' should seek to enhance the overall quality of the production" in terms of time, cost, accuracy and completeness. As further explained, this includes "using iterative and adaptive processes that allow for learning and correction, and, where appropriate, making use of statistically valid metrics in order to monitor progress and obtain valid measures of the accuracy of the effort."[6]

3. Cooperation During the Discovery Process

In 2009, The Sedona Conference® *Cooperation Proclamation* declared that "cooperation by all parties in the discovery process" promotes achievement of the goal of a "just, speedy and inexpensive determination of every action," consistent with the dictates of Fed. R. Civ. P. 1. Among the methods included in the *Cooperation Proclamation* aimed at accomplishing cooperation among counsel are

---

4 *See generally Victor Stanley, Inc. v. Creative Pipe, Inc.,* 250 F.R.D. 251, 261-263 (D. Md. May 29, 2008) (discussing the *Sedona Search Commentary* practice points).

5 The Sedona Conference,® *The Sedona Conference® Commentary on Achieving Quality in the E-Discovery Process*, 10 SEDONA CONF. J. 299 (2009), *available at* http://www.thesedonaconference.org/content/miscFiles/publications_html.

6 *See Pension Committee of the Montreal Pension Plan v. Banc of America,* 2010 WL 93124, *3 (S.D.N.Y. Jan. 15, 2010) (Scheindlin, J.) ("failure to assess the accuracy and validity of selected search terms" constitutes "negligence"); *Victor Stanley, supra,* 250 F.R.D. at 256 (defendants "regrettably vague" on how keywords searched "were developed, how the search was conducted, and what quality controls were employed to assess their reliability and accuracy"); *William A. Gross Construction Associates, Inc. v. American Manufacturers Mutual Insurance Co.,* 2009 WL 724954 (S.D.N.Y. Mar. 19, 2009) ("this Opinion should serve as a wake up call to the Bar … about the need for careful thought, quality control, testing and cooperation with opposing counsel in designing search terms or 'keywords' to be used"); *In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. 650, 660 n.6, 662 (M.D. Fla. 2007) (criticizing defendant's failure to "assure reasonable completeness and quality control" in search for relevant material).

- "Exchanging information on relevant data sources, including those not being searched..."

- "Jointly developing automated search and retrieval methodologies to cull relevant information."[7]

"The Case for Cooperation," published as a supplement to volume 10 of *The Sedona Conference Journal*,® goes on to say:[8]

> "[W]orking cooperatively with opposing counsel to identify a reasonable search protocol, rather than making boilerplate objections to the breadth of a requested protocol or unilaterally selecting the keywords used without disclosure to opposing counsel, may help avoid sanctions or allegations of intentional suppression. Indeed, because knowledge of the producing party's data is usually asymmetrical, it is possible that refusing to 'aid' opposing counsel in designing an appropriate search protocol that the party holding the data knows will produce responsive documents could be tantamount to concealing relevant evidence." [9]

4. Placing "Search Negotiations" in Context: Custodians, Dates, and Scope

In order to appropriately search for responsive and relevant ESI, the responding party will need to identify key custodians familiar with the allegations in the pleadings filed in a case. They are often in the best position to identify what types of information relate to the matter and where they are stored. Information gleaned from these custodians can greatly assist, in addition to the allegations in the pleadings, in determining preservation requirements and whether the information retention schedules would suggest what information is likely to be extant.

B. Search Methods 101: Keywords and Their "Alternatives"

All lawyers are familiar with keyword searching from Westlaw and Lexis searches performed against structured bodies of case law and legislation. However, as the *Sedona Search Commentary* addresses at length, and a growing body of case law recognizes, the exponential growth of ESI coupled with the ambiguities of human language pose profound challenges to constructing efficacious keyword searches for the purpose of finding all or most relevant documents in a given collection. *See Sedona Search Commentary, passim.* Accordingly, any negotiations over search terms must start with the assumption that simple recitation of "keywords," without further refinement, will necessarily end up being both under-inclusive (as there may be relevant documents that fail to contain the string of letters comprising a given keyword), and over-inclusive (in that a potential huge number of false positive, nonrelevant "hits" can be expected with the input of any common term). Parties should recognize the danger in proposing (or accepting a search based on) large numbers of keywords.[10]

---

7 The Sedona Conference,® *Cooperation Proclamation,* 10 SEDONA CONF. J. 331, 332 (2009 Supp.).

8 The Sedona Conference,® *The Case for Cooperation,* 10 SEDONA CONF. J. 339 (2009 Supp.).

9 *See also Dunkin' Donuts Franchised Restaurants, Inc. v. Grand Central Donuts, Inc.,* 2009 WL 1750348 (E.D.N.Y. June 19, 2009) (citing the *Cooperation Proclamation*, parties "are directed to meet and confer on developing a workable search protocol").

10 *See In re Fannie Mae Litigation,* 552 F.3d 814 (D.C. Cir. 2009) (appellate court upholds contempt citation against government agency, where it had failed to meet court-imposed deadlines after agreeing to produce non-privileged documents found responsive to 400 keyword search terms, where the production set consisted of 660,000 recovered documents that needed to be, but could not be, reviewed in time); *see also William A. Gross Construction, supra* (court notes 1,000 proposed search terms put forward by one party, including very generic ones); *Kipperman v. Onex Corporation,* 2008 WL 4372005 (N.D. Ga. Sept. 19, 2008) (defendants' motion for relief from having to review and produce all results from plaintiffs' proposed searches denied, where defendants had failed to timely object to scope of email search both with respect to named end users and search terms).

As the *Sedona Search Commentary* also recognized, due to errors inherent in the transcribed words in texts, keyword searches that are limited to correct spellings, without also including commonly misspelled variants of words, or variations of words using different word stems, are at risk of being incomplete. Counsel should be alert to the need to account for possible variants of words, and consideration should be given to software that employs principles derived from "fuzzy logic."[11]

As the *Sedona Search Commentary* goes on to note, "[l]awyers are beginning to feel more comfortable using alternative search tools to identify potentially relevant" ESI, including language-based and statistical tools that fall under the umbrella of "concept based" methods. United States Magistrate Judge Judge M. Facciola, writing in *Disability Rights Council of Greater Washington, et al. v. Washington Metropolitan Transit Authority,*[12] was the first to suggest in a reported case that parties consider discussing "concept searching" as a possible alternative to keyword searching. Chief Magistrate Judge Paul W. Grimm's *Victor Stanley* opinion goes on to list a number of such methods for parties to generally consider.[13] Although such tools and techniques are coming into greater use, there is as of this writing no reported case law where one or more parties have proposed an alternative search method be employed giving rise to a disagreement that needed court adjudication. Ideally, as such techniques come into their own, they will be subject to utilization and adoption in search protocols memorialized at meet and confers.[14]

A growing number of lawyers in particular are employing statistical clustering techniques that group together "like" documents (documents with similar terms or concepts) into categories. Doing so greatly decreases the amount of time of time needed in manual document review for purposes of responsiveness and privilege.[15] The extent to which a party's decision to employ this type of emerging technique is an appropriate (*e.g.,* nonprivileged) subject within the scope of a negotiated search protocol is an open issue.

C. Approaching the Search Negotiation: Strategies, Models, and Best Practices

While there has been much discussion about the efficacy of search terms, just as there has been with human review, one must begin somewhere, and search terms are commonly used and when used appropriately with a sound methodology, can yield good results. The traditional approach in this area has been for the responding party, after a review of the pleadings and in consultation with key custodians, to draft a list of terms to be searched. Increasingly parties are finding that a sample search can then be run to determine if the terms are likely to be under- or over-inclusive and can then be refined. In addition, a sample of information that has not been included as a result of the search can be reviewed to determine if other terms or alterations should be made.

11 *See Kay Beer Distributing, Inc. v. Energy Brands, Inc.,* 2009 WL 649592 (E.D. Wis. June 10, 2009) (holding that defendant must conduct a search for variants of plaintiff's name).
12 242 F.R.D. 139 (D.D.C. 2007).
13 250 F.R.D. 251 at 259 n.9 (D. Md. May 29, 2008).
14 *See generally United States v. O'Keefe,* 537 F. Supp. 2d 14 (D.D.C. Feb. 18, 2008) (in a criminal case, the court ordered further explanation of whether keyword searches were thorough, citing to authorities arising in civil case law, and suggested that in light of interplay of the sciences of computer technology, statistics, and linguistics, expert testimony may be needed in this complex area).
15 Ronni D. Solomon & Jason R. Baron, *Bake Offs, Demos & Kicking the Tires: A Practical Litigator's Brief Guide to Evaluating Early Case Assessment Software & Search & Review Tools,* Apr. 9, 2007, *available at* http://www.kslaw.com/Library/publication/BakeOffs_Solomon.pdf.

1. The Pros and Cons of Adopting the Requesting Party's Requests With Respect to Keyword Search Terms and Methodologies

A requesting party may have a legitimate, good faith belief that they are sufficiently informed regarding the causes of action at issue and underlying facts so as to be able to propose well-formed search queries, including through the use of keywords. As the propounder of the eventual discovery requests, the requesting party is in the best position to know what it believes are the most salient aspects of the case that are in need of discovery in the first place. To the extent the producing party is willing to allow the requesting party to control all or some of the keyword search, without raising a threshold objection, doing so holds out the possibility of significantly reducing the level of conflict and subsequent motion practice in discovery. An early agreement on search protocols also may be strategically valuable in diminishing the other side's ability to object to terms and methods that have been made subject to prior agreement (especially if they entail large resources or workloads). Cooperation in the form of reaching agreement on search terms ultimately reduces the legal risk in having to undertake new and different searches through large collections of data.

On the negative side, allowing a party with less information and less access to a data set to control the terms of a search often results in proposals for long lists of terms, simply due to unfamiliarity with the universe of "internal" terms that may be used, as well as the understandable desire to capture as much requested information as possible. This can result in the inclusion of considerable unresponsive information that is expensive and burdensome to both parties and slows advancement to reaching the true issues remaining in contention.[16]

2. The Pros and Cons of Allowing the Producing Party to Control the Keyword Search and Methodology

Given the nature of the discovery process, allowing the producing party to control the keyword search and other forms of search, at least in the first instance, makes eminent sense given that the producing party knows (or at least enjoys greater access to) the data, custodians, and internal terminology that will be made subject to the request. The producing party may hold out that upon receipt of properly framed requests for production of documents and ESI, they will undertake to first formulate their search strategy in light of a good faith interpretation of the content of the discovery requests received. If clarification from the requesting party is necessary to narrow the scope of the request, it will be forthcoming. Proceeding in this fashion does not necessarily mean defaulting to a traditional, unilateralist approach in responding to discovery, including the assertion that all efforts behind the scenes to construct a search strategy should be considered "privileged." Rather, the producing party would be representing that they are merely better positioned to begin the process while allowing for transparency with respect to search results at a later stage of the process. Indeed, a disadvantage can occur if there is not sufficient due diligence in the process to arrive at appropriate terms.

16 *See generally Wixon v. Wyndham Resort Development Corp.,* 2009 WL 3075649 (N.D. Cal. Sept. 21, 2009) (broader terms suggested by requesting party held reasonable); *In re Direct Southwest, Inc. Fair Labor Standards Act Litigation,* 2009 WL 2461716 (E.D. La. Aug. 7, 2009) (court accepts requesting parties' search terms over defendants' objections); *Flying J Inc. TCH LLC v. Pilot Travel Centers, LLC,* 2009 WL 1834998 (D. Utah June 25, 2009) (court to rule on search protocol after requiring justification from requesting party for 28 specific terms, subject to any objections lodged by defendant); *Capitol Records v. MP3 Tunes,* 261 F.R.D. 44 (S.D.N.Y. Aug. 13, 2009) (court steps in to adjudicate dispute after court "directed counsel to confer further in an attempt to agree on search terms," and where nine terms were agreed to with 30 remaining in dispute).

As stated in the *Sedona Achieving Quality Commentary*, the first step in constructing any automated search involves the ability "to effectively share and transfer knowledge among counsel and the managing team and those with knowledge of the corpus of ESI that is the subject of discovery. The knowledge gained in this process will be used in the development of one or more search strategies (*e.g.,* Boolean searches, concept searches, metadata filters, language-based approaches using taxonomies and ontologies, statistical clustering techniques, or other proprietary strategies)."

One leading commentator has written extensively on the tension inherent in allowing a requestor to control the keyword search protocol versus what he views as proper reliance on the producing party to re-assert traditional authority in this area, especially in light of the asymmetry in the position of the parties vis-à-vis knowledge of the ESI repository at issue.[17]

3. Adoption of a Phased or Iterative Approach

The rules require the parties to engage in at least one meet and confer; however, the parties have ample opportunity during the discovery process to engage in further discussions. One model for phased interactions is as follows:[18]

> *Step 1. The parties meet and confer on the nature of each others' computer hardware and software applications. Proposals are exchanged on the scope of search obligations, in terms of databases and applications to be searched, what active and possibly legacy media, key custodians, time periods. Additionally, keywords are proposed along with any other more sophisticated Boolean or concept search methods. A timetable for conducting searches after the propounding of discovery requests is agreed to.*
>
> *Step 2. In the interval between meet and confers, parties conduct searches in accordance with prior representations and the actual wording of discovery requests. In doing so they may utilize sampling techniques, and estimates are gathered on the volume of data or "hits" made subject to search.*
>
> *Step 3. The parties interact further in describing the result of initial searches and preliminary results. If the parties have agreed to a Rule 502 rubric, the parties may elect to share documents found to be potentially responsive. Search terms and protocols are adjusted and search methods are tuned or adjusted for the purpose of conducting more narrow, focused searches.*
>
> *Step 4. The parties may elect to continue iteratively until a mutually agreed time or cap on numbers of responsive documents is reached.*

As noted, a variation on this approach is to presume in Step 1 that the responding party or parties in control of the data to be searched take the lead in first conducting a search under unilaterally arrived at keywords and other search methods. Doing so does not in theory alter further following remaining steps 2 through 4 *supra*. The efficacy of

---

17 Ralph Losey, *Child's Game of Go Fish Is A Poor Model for E-discovery Search*, Oct. 4, 2009, *available at* http://e-discoveryteam.com/2009/10/04/childs-game-of-go-fish-is-a-poor-model-for-e-discovery-search/; *see also Spieker v. Quest Cherokee,* 2008 WL 4758604 (D. Kan. Oct. 30, 2008) ("Since the documents were created, stored, and/or maintained [by defendant], defendant is in the better position to develop the most appropriate list of search terms capable of producing the requested documents," and suggesting that defendants should modify plaintiff's proposed search terms if not specific enough).

18 George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?* 13 RICH. J.L. & TECH. 10 (2007), *available at* http://law.richmond.edu/jolt/v13i3/article10.pdf.

engaging in multiple further "iterations" with opposing counsel to refine terms diminishes over time, and must be judged against available resources and discovery deadlines.[19]

4. Use of Sampling

The producing party should anticipate that regardless of what search protocol or method the parties agree upon, there will also be an expectation that the producing party has engaged in some degree of sampling to assure accurate and complete results. (The extent of the effort undertaken to sample should be proportional to the case's overall complexity, taking into account the degree of risk and what is at stake.) The *Sedona Achieving Quality Commentary* describes in greater detail how to conduct an automated search process utilizing statistical sampling:

> Once the responsive data set has been [arrived at or] characterized, a random sample of categorized material is chosen and reviewers will review this small, but statistically significant sample. This random sample will contain both responsive and unresponsive material, and reviewers classify these documents as they normally would under a manual review. The results of this classification are then compared to the results reached by the chosen categorization method(s). When there is a difference between the determination made by the human reviewer and the categorization method, the legal team reviews the document and decides which is correct. Adjustments are then made to the search strategy. Sometimes the differences require modifications so that a particular type of document is filtered in the future.[20]

5. Documentation and Defensibility

At every stage of the negotiation process, both the responding and requesting party should be well prepared in advance to explain the basis of the search strategies being proposed and the rationale as to why they are believed to yield or to have yielded quality results. At a basic level, counsel and/or any witness to be put forward must know the scope of the search conducted in terms of whose files and which files were searched, what people were told about how to go about conducting the search, and how the search was supervised. Additionally, rigorous documentation of the process is key to defensibility and to defining consistent repeatable additional searches, regardless of whether such documentation is shared with an opposing party. As noted in Practice Point 7 of the *Sedona Search Commentary*, "[t]his explanation may best come from a technical "IT" expert, a statistician, or an expert in search and retrieval technology. Counsel must be prepared to answer questions, and indeed, to prove the reasonableness and good faith of their methods." [21]

19 Feng C. Zhao, Douglas W. Oard & Jason R. Baron, *Improving Search Effectiveness in the Legal Discovery Process Using Relevance Feedback,* DESI III Conference, Barcelona, June 8, 2009, *available at* http://www.law.pitt.edu/DESI3_Workshop/DESI_III_papers.htm; *see generally Ameriwood Industries, Inc. v. Liberman,* 2007 WL 685623 (E.D. Mo. Feb. 23, 2007) (court orders expert report with number of "hits" based on negotiated search terms, with expectation that parties will continue to meet and confer to refine search based on false positives); *ClearOne Communications, Inc. v. Chiang*, 2008 WL 920336 (D. Utah Apr. 1, 2008) (court adjudicates dispute over conjunctive versus disjunctive operators between search terms, urging parties that further refinement of terms is possible subject to additional negotiations).

20 8 SEDONA CONF. J. 189 (2007).

21 *See generally, Victor Stanley,* 250 F.R.D. 251.