

## Once is Not Enough: The Case for Using an Iterative Approach to Choosing and Applying Selection Criteria in Discovery

By Gene Eames, David J. Kessler and Andrea L. D'Ambra

Culling using selection criteria is used to manage otherwise overwhelming volume of data in litigation discovery.<sup>1</sup> Courts and experts are endorsing the emerging best practice of iterative measurement of selection results, borrowing from established and effective data management practices outside of the litigation setting. By insisting on this practice, in-house counsel and their advisers can both contain the cost of discovery and reduce or eliminate the risk of challenge to the search and selection methodology.

The use of search terms and other selection criteria (*e.g.*, date restrictions or other filters) by parties in litigation at each and every stage of discovery, from preservation to collection to privilege review, is commonplace. The courts and other experts recognize this practice as a necessary alternative to the impossible task of reviewing every possible document in the overwhelmingly large collections of electronic data subject to consideration. And, over time, the process for choosing and applying search terms has evolved as parties, the bench and the bar have looked for better and more cost effective methods to handle discovery. But as the use of selection criteria has become an indispensable tool in the e-discovery toolkit, its application has come under increasing scrutiny by opposing parties and the courts in order to ensure that appropriate search criteria are being used and that responsive materials are not systematically missed in the rote application of search terms. In the last couple of years, there have been several high profile e-discovery cases that dealt with the choice of search terms and their application.<sup>2</sup>

---

<sup>1</sup> See Practice Point 1 *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, August 2007 (“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.”)

<sup>2</sup> *Asarco v. EPA*, 2009 WL 1138830 (D.D.C. Apr. 28, 2009) (requiring an iterative search process where plaintiff demonstrated that the original search failed to yield clearly responsive documents); *William Gross Cons. Assoc. v. Amer. Manuf. Mut. Ins. Co.*, 256 F.R.D. 134 (S.D.N.Y. Mar. 19, 2009) (holding that parties must carefully craft appropriate keywords with input from ESI custodians and final search protocol must be quality control tested); *In re Fannie Mae Sec. Litig.*, 552 F.3d 814 (D.C. Cir. Jan. 6, 2009) (denying protective order where non-party had stipulated to defendant’s search terms); *Kipperman v. Onex Corp.*, 2008 WL 4372005 (N.D. Ga. Sept. 19, 2008) (denying a protective order where defendants had failed to previously narrow search terms at court’s suggestion); *Victor Stanley v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. May 29, 2008) (holding that defendant waived its privilege because it failed to show the terms used to cull for privilege were reasonable); *ClearOne Comm., Inc. v. Chiang*, 2008 WL 920336 (D. Utah Apr. 1, 2008) (holding that parties should refine search terms based upon a review of material both returned and not returned by the search); *U.S. v. O’Keefe*, 537 F. Supp. 2d 14 (D.D.C. Feb. 18, 2008) (requiring expert testimony to prove the insufficiency of search terms); *Equity Analytics, LLC v. Lundin*, 2008 U.S. Dist. LEXIS 17407 (D.D.C. Mar. 7, 2008) (requiring an expert report to demonstrate why proposed limitations on search protocol would unreasonably limit discovery); *Qualcomm Inc. v. Broadcom Corp.*, 2008 WL 66932 (S.D. Cal. Jan. 7, 2008) (levying sanctions where plaintiff, among other things, failed to conduct a reasonable search for responsive documents).

As these cases show, there can be significant risks – of both increased discovery costs and various discovery sanctions – if a litigant improperly chooses and applies selection criteria.

The good news is that a reasonable, defensible, best practice approach to using selection criteria and cost-effective discovery are not mutually exclusive. An iterative approach to choosing, testing and applying selection criteria will not only be easier to defend from attack by opposing parties, but in many cases will reduce the overall cost of discovery by eliminating more irrelevant documents from processing and review. Corporate counsel don't have to choose between containing discovery costs and increasing the defensibility of their process – with the iterative approach to selection criteria, they can do both.

### **Why Search at All?**

The goal for producing parties is to carve from the overall universe of electronic documents a relevant subset of documents for review and ultimate production in a legally defensible and reasonable manner. Litigants faced with electronic document requests quickly realize that there is an impossible volume of data that must be assessed. In most cases, the entirety of electronic documents cannot be produced without some level of selection and review. Neither the producing party nor the requesting party would be able to handle the volume, the effort might not be proportionate to the value of the matter, and the production of non-responsive, unrelated business information might be harmful to producing parties.

This task is usually accomplished by selecting documents via some search mechanism. This is not new. From the beginning of discovery, counsel and parties have been making choices of where and what to search for responsive documents. What text searchable electronic documents have done is abstracted the process.<sup>3</sup> A party may choose to look at one custodian's files over another because they are more likely to have responsive documents. A party may not open a filing cabinet or box because the label indicates it does not have any documents of value -- a reasonable decision. However, applying search terms to a large volume of data does not necessarily have that same effect – especially if the party does not know the effectiveness of those criteria.

---

<sup>3</sup> For purposes of this article we have limited ourselves to a discussion of text searchable electronic data since a traditional e-discovery search is intended to be applied to this class of data. These data include the common electronic documents with which most of us are familiar – e-mail communications, and user-generated language-based documents created in Word, PowerPoint, Excel, etc. – sometimes referred to as “unstructured data” because the applications used to create these files have no restrictions placed on them as to content. A user can include any subject matter, include any number of different topics and is not limited to any particular vocabulary, spelling, grammar or general writing style. The language contained in text documents includes nuances, ambiguities and inaccuracies. Most documents found in an e-discovery collection have not been subjected to any editorial review, nor have they had restrictions imposed on vocabulary usage. These factors compound the difficulties in effectively applying search strategy to identify documents that are poorly understood or completely unknown prior to review. The nature of unstructured data makes text searching an inexact science.

## Measuring Effectiveness of Searches

While the courts and other experts have recognized searches as a necessary tool in discovery, they have also recognized that text searching is a method with inherent difficulties. In the best of circumstances, search terms are both over- and under-inclusive. People may use a term out of context that creates a pool of non-responsive documents; likewise, employees may use jargon unknown to counsel (leading to responsive documents not being hit by search terms). Even the use of date ranges, which are seemingly objective and non-controversial, can cause errors – what happens when an e-document is misdated or not dated at all? Just as counsel take reasonable steps to ensure quality control on review calls made by human reviewers in discovery, they should consider undertaking quality control steps to ensure their selection criteria effectively recall relevant documents without collecting too many irrelevant documents.

Two important concepts are related to measuring the efficacy of searches: recall and precision. Recall refers to the comprehensiveness of a search – has it identified all of the ultimately responsive documents you sought? Precision refers to the accuracy of the search – of the documents returned by your search, how many were ultimately deemed responsive, and how many were false positives or false hits?

An effective search balances both good recall and good precision, which is sometimes referred to as “targeting” a result set. Judge Grimm refers to the issues of precision and recall in the *Stanley* decision when he notes that “[c]ommon sense suggests that even a properly designed and executed keyword search may prove to be over-inclusive or under-inclusive...”<sup>4</sup> It is possible that a search result can have poor recall (being under-inclusive), or poor precision (being over-inclusive), or both.

The ultimate test of search effectiveness looks at responsiveness rates in review sets and ensures that the searches have not left responsive documents out of review altogether. A high percentage of ultimately responsive documents in the total universe of documents reviewed indicates good precision – not too many non-responsive documents were returned by the search. A low responsiveness rate would indicate poor precision, meaning many false hits were returned.

Looking at the responsiveness rates of review sets alone cannot help determine whether good recall was achieved, however. Good recall can only be determined by looking at the total universe of documents, or samples of those documents, and not just the sets actually returned by the searches. Ensuring comprehensive searches typically can be accomplished by selective expansion and broadening of search criteria, and by sampling what the searches are returning and not returning. While the case related to a privilege determination and not to responsiveness, Judge Grimm expressly refers to this practice in the *Stanley* case, “The only prudent way to test the reliability of the keyword search is to perform some appropriate sampling of the documents determined to be privileged and

---

<sup>4</sup> See *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008).

those determined not to be in order to arrive at a comfort level that the categories are neither over-inclusive nor under-inclusive.”<sup>5</sup>

A number of undesirable outcomes can occur if precision and recall are out of balance. Good recall accompanied by poor precision can result in an extremely over-inclusive review set, basically rendering the search attempt useless. This can occur when overly broad un-tested terms are used in the search, or when automated search tools – automated methods for expanding searches by employing a thesaurus, ontology, taxonomy, derived-related terms and concepts, alternate spellings, homonyms, and so on – expand searches in a nonselective manner. When these expansions are employed without human interaction and selective implementation, an unintended explosion of the search result can occur. The resulting flood of documents may reduce the likelihood of missing relevant documents by increasing recall, but it also renders the search useless by attaining such poor precision that virtually no targeting has occurred, as evidenced by responsiveness rates dropping precipitously low. Even excessively broad search results do not guarantee perfect recall, as there can still be responsive documents in the small number of documents not returned by the overbroad searches. Sampling documents not returned by overly broad searches is still prudent and necessary.

A search outcome with good precision but poor recall is easy to imagine. An extremely narrow search might achieve extremely high percentages of ultimately responsive documents in review – good precision – but large numbers of ultimately responsive documents might be missed by the search and left entirely out of review – poor recall. This scenario may be cost effective, but ultimately incomplete, possibly noncompliant and increases the risk of effective defensibility challenges. A search with good recall and poor precision lessens the likelihood of a defensibility challenge, but it may be anything but reasonable, practical or cost effective. These types of results can be caused by a lack of effort to determine actual effectiveness of search terms, instead relying on an assumption that searches will return an expected result.<sup>6</sup>

The most likely outcome when parties employ the commonly accepted approach of brainstorming searches and applying them in an untested manner is both poor recall and poor precision. But this isn't even the biggest problem with this method of determining and applying search criteria. The biggest problem is – determining that there is a problem. Looking at the records reviewed, if the search resulted in a large volume of records but very few responsive documents – is this because there are few responsive documents or because the search used the wrong terms or the wrong location? A search that results in a large number of responsive documents may be the result of a search that was targeted correctly – or because any net in so rich a fishery of responsive material would have captured a great many documents and perhaps many more were left behind in the data sea.

---

<sup>5</sup> *Victor Stanley*, 250 F.R.D. at 257.

<sup>6</sup> See, for example, *In re Fannie Mae Sec. Litig.*, 552 F.3d 814 (D.C. Cir. 2009), where not only a very expensive review was encountered, but contempt citations resulted when deadlines were missed because of the large volume of review.

The goal of any e-discovery selection criteria is a search result where both good recall and good precision are achieved – a review set with a high responsiveness rate, accompanied by a reasonable assurance that everything that was supposed to be reviewed has in fact been included for review.

### **Traditional Approaches to E-Discovery Search**

Historically, there have been two common approaches to the application of searches in an e-discovery project, namely: a brainstormed list of search terms; or an agreed upon set of search terms. In reality, one is an outcropping of the other as they are both simply brainstormed lists of search terms, the latter being brainstormed by a larger group representing multiple parties, and the former being determined by one party only. The brainstorming practice involves a group of individuals consulting and ultimately guessing (in ostensibly a reasonably educated way) about what search terms should be used to retrieve candidate documents for responsiveness. The problem is that information retrieval experts have repeatedly found that because of the difficulties in accurately predicting which search terms will bring about good precision and good recall, a brainstormed approach to determining search criteria may not prove effective in text searching scenarios.

While brainstormed lists determined by one party have other weaknesses, their primary fault is that they are prone to challenge in hindsight. Even the best searches will miss certain records that in a perfect world would be selected and will include irrelevant documents that one does not want to process or review, but searches based on brainstormed terms are particularly hard to defend – why were some terms chosen and not others?<sup>7</sup> This is never truer than when an opposing party is asking specific questions about terms found in the responsive documents (or in deposition) that were not considered when the initial documents were collected. This recall failure (otherwise known as too many false negatives) can be an easy target for opposing parties – they only bring the challenges after some discovery has already been conducted, so they have better information about the content of the documents and possible appropriate terms. In essence they have sampled (*i.e.* tested) the responsive data and are challenging whether the search was complete. While the *Qualcomm v. Broadcom* case has become infamous for a number of e-discovery and other failures, one of the most basic problems for which the court chastised the Qualcomm lawyers was not using the most obvious search terms to identify and collect obviously relevant material.<sup>8</sup> Even in cases where lawyers and parties brainstorm selection criteria in good faith, their imperfect knowledge as to how people actually created the documents and the terminology unnecessarily limits their results.

---

<sup>7</sup> See, e.g., *Capitol Records, Inc. v. MP3Tunes, LLC*, 261 F.R.D. 44, 48 (S.D.N.Y. 2009) (court chastised defendant’s counsel for asking his client to search using only a single word).

<sup>8</sup> *Qualcomm Inc.*, 2008 WL 66932 at \*10 (“The fact that Qualcomm did not perform these basic searches at any time before the completion of trial indicates that Qualcomm intentionally withheld the documents.”).

Agreed-upon search lists do have the advantage that they are harder to challenge. But while those in favor of agreed-upon terms might suggest that once agreement is reached on search strategy, all problems associated with the ineffective search are avoided, this is not the case. First, using agreed-upon selection criteria can handcuff a responding party and make it very difficult for it to research its own systems to find potentially valuable information to support its own case. Second, it can cause the responding party to spend an inordinate amount of money processing and reviewing irrelevant documents.<sup>9</sup> Fewer disputes may arise about agreed-upon search terms, not because the searches are significantly better when one party chooses the terms, but because once a party agrees to using a certain set of terms, that party loses its right to complain – which may explain judges’ support for early negotiation and agreement on selection criteria, even when parties have not yet collected the data, much less examined it. Judges prefer to minimize disputes and conflict over discovery, and they may be unaware of the specific costs caused by the agreements or the actual effectiveness of the agreed-upon terms.

Of course, even agreed-upon search terms are sometimes questioned after the fact when one side or the other discovers the outcome of the searches to be different than what had been expected, diminishing the perceived value of the agreement in the first instance. In the *Kipperman* decision, the court had ruled that the defendant had implicitly agreed to plaintiff’s brainstormed search terms by neglecting to modify the searches when given the opportunity.<sup>10</sup> The court let what turned out to be an overbroad search stand. If the defendant had put some effort into testing the effectiveness of the searches prior to review, there may have been a different outcome. A similar outcome arose in the *Fannie Mae* case where the producing party had been given the opportunity to participate in the determination of “appropriate search terms,” but failed to do so.<sup>11</sup>

### **Iterative Approach to E-Discovery Search**

There is a better way, and to find it one should look to the industry experts (Sedona, Legal TREC and EDRM), as well as to recent court decisions. These experts in turn have been looking to the information retrieval world beyond e-discovery where these search inadequacies have long been identified, and where experts have been addressing the problems for quite some time. The challenges remain, but an awareness of them allows for planning and choosing methodologies we might use to meet those challenges. Importantly, the standard is *reasonableness*, not perfection,<sup>12</sup> and the thoughtful use of search tools. The Sedona guidelines state that “Lawyers must recognize that, just as important as utilizing the automated tools, is tuning the *process* in and by which a legal team uses such tools, including a close involvement of lead counsel. This may require an

---

<sup>9</sup> See *In re Fannie Mae*, 552 F.3d 814

<sup>10</sup> See *Kipperman v. Onex*, 2008 WL 4372005 (N.D. Ga. 2008).

<sup>11</sup> See *In re Fannie Mae Sec. Litig.*, 552 F.3d 814 (D.C. Cir. 2009). The court let the overbroad search result stand which resulted in an extremely expensive review. *Id.* at \*8. Notably, in both of these instances parties were given the opportunity to develop more effective searches, but did not do so.

<sup>12</sup> *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, August 2007 at 214.

iterative process which importantly utilizes feedback and learning as tools, and allows for measurement of results.”<sup>13</sup>

Whatever search technology one chooses to employ, it is important that one does so in the context of an iterative process. The process should include how the tools are to be employed; what human input and decision-making is required; iterative sampling and measurement of results; and validation that the technology worked as expected. The process must allow counsel to use the findings and alter the terms – there is no point in testing the search terms if it is prohibitively expensive to change them. To test and compensate for poor recall or precision, the process must iterate through multiple attempts at conducting the search correctly. By testing the data (both what is selected and what is not selected), one can mitigate the risk of systematically missing data and create documentation regarding the reasonableness of the process, while at the same time reduce the amount of money wasted on processing and reviewing irrelevant documents. In fact, in larger reviews, the iterative approach can reduce enough data that it will pay for itself many times over.

Search process and verification efforts should be documented – it is one of the benefits for undertaking the process in the first place. The documentation should include explanation of the general process within which the search technology was applied, including all steps taken to measure and validate results, and all corresponding changes made to search strategies based upon those results. This should be maintained in anticipation of explaining to the court if requested.<sup>14</sup> This documentation will serve as evidence if it becomes necessary to assert that search terms are or are not appropriate, and the resulting empirical evidence may be bolstered with expert testimony.<sup>15</sup>

### **Opportunities for Applying E-Discovery Search**

Search technology can also be employed effectively in other instances besides selecting a review set. Criteria can be used as a means to classify or prioritize documents – a privilege screen is an example of this kind of use. A review set has been determined and potentially privileged documents need to be carved out from documents deemed potentially responsive, but which are not likely to be privileged. Privilege classification criteria can be used to perform this division. And there are other reasons to run classification criteria to group documents by subject matter for review efficiency.

Generally, classification criteria have a higher tolerance for inaccuracy because when intended solely to provide review efficiency, and not to determine inclusion or exclusion from review entirely, the risks are lower – a mistake changes when a document is

---

<sup>13</sup> See *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, August 2007 at 199.

<sup>14</sup> See Judge Facciola’s decisions in *O’Keefe* or *Equity Analytics* where he in effect says to prove it.

<sup>15</sup> See *United States v. O’Keefe*, 537 F. Supp. 2d 14 (D.D.C. 2008); see also *Analytics, LLC v. Lundin*, 2008 U.S. Dist. LEXIS 17407 (D.D.C. 2008).

reviewed, not whether it is reviewed at all.<sup>16</sup> Counsel should remember when using classification criteria to undertake privilege review that if privilege materials are produced inadvertently, the party may be obligated to establish that it used reasonable precautions,<sup>17</sup> and a party may need to demonstrate that it employed a reasonable method for settling upon certain classification criteria. This is very different than the use of selection criteria in collection or processing, where the objecting party generally has that obligation.<sup>18</sup>

Criteria can also be employed in a quality control process. For example, to help ensure consistency of review calls and markings prior to final production, criteria can be run across production sets to try to identify anomalies.

In an effort to reduce costs by reducing volumes of data, it is becoming more and more common for parties to apply search terms earlier and earlier in the discovery process, including in preservation and collection filters. Of course, in many instances this practice is simply the same old wine in a new bottle, and in fact processing has simply become portable and easier to use behind a company's firewall. Regardless of where in the EDRM lifecycle data selection is perceived to be occurring, practitioners should be careful when applying criteria. Be mindful that the realities of searching data still hold and in fact the consequences may become potentially more severe at the earliest stages of discovery. Remember that inadequate selection criteria that exclude documents from review have determined those documents to be non-responsive, and possibly incorrectly so. At the preservation and collection phase, correcting these mistakes long after the fact can be quite expensive. If necessary, take a conservative approach and err on the side of over-inclusion until a more controlled environment presents itself where the selection criteria can be tested and validated using iterative samples. Counsel are wise to heed the tone of Judge Peck's decision in the *Gross* case: "This opinion should serve as 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...it appears that the message has not reached many members of our Bar."<sup>19</sup>

---

<sup>16</sup> There are some possible exceptions to this general rule of tolerance. If classifications are to be used for bulk marking, then the impact of false hits or missed hits becomes more significant and consequential. In addition, in the special case of privilege classification the effectiveness and reasonableness of the classification attempt may play an important role in determining whether sufficient steps were taken to avoid inadvertent production of privileged documents. There is a potential for waiver of such privilege if a court determines that a privileged set of documents were produced, and further determines that the steps taken to avoid such an occurrence were not adequate. See *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008); see also *Rhoads Indus., Inc. v. Bldg. Materials Corp. of Am.*, 254 F.R.D. 216 (E.D. Pa. 2008). To avoid such an outcome, the iterative steps used in testing, measurement and validation of selection criteria development may also be incorporated in the development of classification criteria.

<sup>17</sup> See FRE 502; *Victor Stanley*, 250 F.R.D. 251.

<sup>18</sup> See *The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process*, May 2009 at 19; The Sedona Conference Best Practices, Recommendations and Principles for Addressing Electronic Document Production, Principle 7 ("The requesting party has the burden on a motion to compel to show that the responding party's steps to preserve and produce relevant electronic data and documents were inadequate.").

<sup>19</sup> See *William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134 (S.D.N.Y. 2009).



## **CONCLUSION**

Selection criteria are here to stay and, in fact, have always been with us. Increased scrutiny by courts and opposing parties require counsel to ensure the search techniques they use are reasonable. Counsel should look to iterative search and selection procedures to better choose their search terms and control costs. Selecting more refined selection criteria is not only about being more defensible and finding more responsive records; it is also about discarding irrelevant material early in the process and containing costs. Put differently, fewer search terms (and less time thinking, analyzing, and refining them) does not necessarily lead to smaller search results and lower cost.