WHITE PAPER

Guidelines for Defensible Data Reduction in eDiscovery

An Overview of the Law of Search Methodology

Using technology to search, review, and analyze electronically-stored information (ESI) has become critical in litigation for a number of reasons. For one, manual search and review of all of the ESI in a typical complex litigation matter is usually too costly and time-consuming to be feasible.

Moreover, manual review is not necessarily more accurate or reliable than other methods, so there is no reason to assume manual review is a best practice for all cases. In an often-cited 1985 study, D.C. Blair and M.E. Maron showed that in a 40,000 document case, search terms found only 20 percent of relevant documents, though lawyers estimated that they would find 75 percent.[1]

Thus, litigants are increasingly turning to technology for assistance in culling, searching, and reviewing massive data sets to find relevant information and avoid producing privileged information. A wide array of tools is available to help them, including keyword searching, concept searching, statistical search techniques, clustering, taxonomic approaches, and predictive coding.[2]

As with most technologies, however, there are better and worse ways to use the technology for the task at hand. This paper will provide guidelines for appropriate and defensible use of technology in litigation. It is critical that the methodology employed be defensible because the penalties for misuse can be severe, ranging from steep monetary sanctions[3] to waiver of attorney-client privilege.[4] Recent surveys of ediscovery case law suggest that while courts have continued to grant motions for sanctions at approximately the same rate, the frequency with which sanctions are awarded is increasing because they are being sought more often.[5] In some cases, courts have even crafted search methodologies themselves when litigants cannot agree, which can be risky and lead to results that are less than ideal.[6]

When it comes to the use of technology in early case analysis and review, clear standards of defensibility are still emerging. However, some general guidelines can be gleaned from the existing case law. Although the cases discussed here focus on keyword searching, the guidelines that emerge are general and can be applied to other methodologies.[7]

Reasonableness is the Touchstone

In the context of electronic search and retrieval, defensibility implicates two basic standards. Both boil down to a standard of reasonableness.

First, Federal Rule of Civil Procedure 26(g)(1)(A) requires that an attorney responding to a discovery request attest that "to the best of the person's knowledge, information, and belief formed after a reasonable inquiry … with respect to a disclosure, it is complete and correct as of the time it is made." *See also Zubulake v. UBS Warburg, LLC*, 229 F.R.D. 422 (S.D.N.Y. 2004) ("Zubulake V") (holding that "counsel and client must take some reasonable steps to see that sources of relevant information are located"); *In re Seroquel Prods. Liability Litig.*, 244 F.R.D. 650, 663 (M.D. Fla. 2007) (holding that attorneys have a responsibility at the outset of the litigation to "take affirmative steps to monitor compliance so that all sources of discoverable information are identified and searched"); *Eurand, Inc. v.*

Mylan Pharms., Inc., 266 F.R.D. 79 (D. Del. 2010) ("In evaluating whether search terms or search methods employed to carry out the search were appropriate, the court applies a 'reasonableness test to determine the "adequacy" of search methodology.'").

Second, Federal Rule of Evidence 502(b), enacted in 2008, provides that an inadvertent disclosure of material covered by attorney-client privilege or work-product protection does not operate as a waiver if the holder of the privilege or protection "took reasonable steps to prevent disclosure," and "promptly took reasonable steps to rectify the error."

It should also be noted that the Federal Rules of Civil Procedure require that parties confer "as soon as practicable" to discuss discovery, including ediscovery: "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced." Fed. R. Civ. P. 26(f). This is significant here because, as explained below, cooperation and planning are crucial to using technology defensibly. Being prepared and informed before Rule 26(f) conferences can greatly impact the defensibility of decisions made down the road.

An Overview of the Law

Courts have concluded that the use of technology can meet the standard of reasonableness and, as a general matter, have fully accepted the use of technologies such as keyword searching as a substitute for manual search and review. *See, e.g., Treppel v. Biovail Corp.*, 233 F.R.D. 363 (S.D.N.Y. 2006) (where plaintiff would not agree to defendant's proposed keyword search methodology for locating documents responsive to plaintiff's discovery requests but had refused to suggest alternative methodologies, ordering that defendant should proceed with its proposed search); *Victor Stanley*, 250 F.R.D. at 260 ("[K]eyword searches have long been recognized as appropriate and helpful for ESI search and retrieval…").

However, courts have shied away from endorsing specific technologies over others.[8] That may change as ediscovery decisions proliferate and courts become more knowledgeable about the issues involved. On the other hand, courts may continue to be reluctant to provide specific guidelines given the rate of change in technology and the fact-specific nature of typical ediscovery disputes.

Nonetheless, it is possible to glean some general guidelines from the existing case law. Unlike in other areas of ediscovery such as the duty of preservation, where courts are widely divided on standards of defensibility, courts have been generally consistent in their treatment of electronic search and retrieval technologies.

In the key cases discussed below, courts address the use of electronic search and retrieval technologies and, in doing so, provide some advice on how to use technologies defensibly.

VICTOR STANLEY, INC. V. CREATIVE PIPE, INC.

In *Victor Stanley*,[9] defendants accidentally produced privileged documents. The court held that they had waived attorney-client privilege and work-product protection because they failed to demonstrate that their ediscovery search methodology reasonably could have prevented the inadvertent production of privileged documents.

The parties had agreed on search terms that addressed responsiveness and defendants used those search terms to retrieve ESI. The defendants indicated that they planned to use keyword searches to locate privileged documents, and requested that the court approve a clawback agreement to address any inadvertent disclosures. However, before the court approved the agreement, the defendants changed their minds and said that they would instead do an individualized privilege review.

What they actually did, however, was perform keyword searches of those documents that were text-searchable, and perform a manual review of the titles of those documents that were not text-searchable. The documents flagged by each of these searches were then reviewed, but no sampling or review was done of the documents not retrieved by those searches.

Defendants' first production contained 165 privileged documents, and plaintiffs contended that defendants had waived privilege. Magistrate Judge Paul Grimm agreed, noting that defendants had the burden to

demonstrate that their search methodology was reasonable because of the "well-known limitations and risks associated with [keyword searches]."

Judge Grimm outlined some general guidelines for defensibility, including holding that defendants should have conducted sampling of the documents not recalled by the privilege searches:

> Selection of the appropriate search and information retrieval technologies requires careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected must be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented.

IN RE SEROQUEL PRODUCTS LIABILITY LITIGATION

In *Seroquel*,[10] plaintiffs sought sanctions based on defendant's alleged discovery failures, including a belated and incomplete production rife with technical errors and in non-searchable formats.

Among other things, plaintiffs contended that the defendant performed an "unreasonable key word search" because of its failure to include such things as common misspellings, singular forms of words, or the generic name for the drug at issue in the lawsuit. The court agreed that sanctions were warranted, holding that "the key word search was plainly inadequate. Attachments, including non verbal files, were not provided. Relevant emails were omitted. [Defendant's] deduplication method remains mysterious." The court continued:

> [Defendant] purported to embrace the requirements of Rule 26 and the Sedona Principles. However, the reality was to the contrary. For example, while keyword searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process. Rather than working with Plaintiffs from the outset to reach agreement on appropriate and comprehensive search terms and methods, [Defendant] undertook the task in secret. Common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness. If [Defendant] took such steps, it has not identified or validated them.

The court was also troubled by the facts that the defendant shielded its third-party technical contractor from contact with plaintiffs instead of allowing cooperation; that defendant's attorney could not testify as to how the keyword search was developed; and that defendant never discussed with plaintiffs which search terms to use.

WILLIAM A. GROSS CONSTR. ASSOCS., INC. V. AM. MFRS. MUT. INS. CO.

Unlike *Victor Stanley and In Re Seroquel, Gross*[11] did not involve sanctions or a privilege waiver. In Gross, Magistrate Judge Andrew Peck found himself in the position of crafting search terms to resolve a discovery dispute over a party's search requests to a non-party. The non-party protested that the proposed search terms were blatantly overinclusive but did not offer assistance in narrowing the proposed searches. Thus, the court was called on to resolve the dispute. Judge Peck lamented: "This case is just the latest example of lawyers designing keyword searches in the dark, by the seat of the pants, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails."

Then, the Judge Peck took the opportunity to issue a "wake-up call" to lawyers "about the need for careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms or 'keywords' to be used to produce emails or other electronically stored information ("ESI")":

> 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.

Putting it Together: Guidelines and Tips for Making Defensible Decisions

From these cases and others, some general guidelines for using technology defensibly can be determined. Though most cases that discuss the use of technology in ediscovery have focused on keyword searching, there is no reason to believe that these general guidelines will not apply to more advanced search methodologies.

1. WHEN SELECTING A SEARCH METHODOLOGY, USE CARE AND PLANNING, AND BE INFORMED.

Victor Stanley, *Gross*, and *In re Seroquel* make clear that decisions about using technology in ediscovery, such as decisions about keywords, are far more likely to be defensible when they are the informed product of careful thought and planning.

This means getting input from those who know the documents, including key custodians, and from those who understand the company's IT systems and the ediscovery technologies being used. Assistance from search methodology experts and experts in particular ediscovery tools can also be invaluable. Even if you understand the documents and the words or concepts that you should be looking for, it is essential to understand *how* to find them with the ediscovery technology you've decided to use.

2. TEST, TEST, AND TEST AGAIN.

Even the best-planned search methodology—whether to locate responsive documents or to identify privileged documents—might not be defensible without quality control procedures, such as sampling of documents that are not retrieved by searches.

Victor Stanley, *Gross*, and *In re Seroquel* all explicitly make the point that a chosen search methodology must be tested for quality assurance. *See also Ingersoll v. Farmland Foods, Inc.*, 2011 WL 1131129, No. 10-6046-CV-SJ-FJG (W.D. Mo. Mar. 28, 2011) (where parties could not agree on a protocol for producing ESI, denying plaintiff's motion to compel production immediately without any sampling and instead ordering defendant to proceed with its proposed sampling procedure to evaluate the appropriateness of search terms before producing documents); *Mt. Hawley Ins. Co.v. Felman Prod., Inc.*, No. 3:09-CV-00481, 2010 WL1990555 (S.D. W. Va. May 18, 2010) (finding that the precautions taken to prevent disclosures were not reasonable, and that therefore privilege was waived, where, among other issues, the party claiming privilege had failed to test keyword searches by appropriate sampling).

3. COOPERATION IS KEY.

In re Seroquel and *Gross* both emphasize the importance of cooperation between opposing parties when developing and implementing a search methodology and deciding on search terms, and revised Fed. R. Civ. P. 26(f) mandates that parties confer on ediscovery issues.

In 2008, the Sedona Conference, a non-profit legal research and educational institute considered a leading authority on ediscovery, issued a "Cooperation Proclamation," calling for "open and forthright information sharing, dialogue (internal and external), training, and the development of practical tools to facilitate cooperative, collaborative, transparent discovery."[12] The Cooperation Proclamation has been endorsed in a number of ediscovery cases, including *Gross*.[13]

Of course, it may not always be possible to cooperate with the opposing party. The opposing party may be unwilling to cooperate, or in some cases, cooperation may not be aligned with your strategy. But, where possible, genuine efforts at cooperation with opposing seem to go a long way toward convincing a judge of the reasonableness and defensibility of a party's use of technology. And, with the opposing party's agreement on search terms and methodologies, later challenges will be less likely.

4. WHERE POSSIBLE, COMPLY WITH SEDONA CONFERENCE BEST PRACTICES.

The Sedona Conference has published several useful resources for ediscovery, including The Sedona Conference Best Practices Commentary on The Use

of Search & Retrieval Methods in E-Discovery (August, 2007). This and other publications are available at the website listed at footnote 14.

A number of ediscovery cases, including *Victor Stanley*, *Gross*, *Treppel v. Biovail*, and *In re Seroquel*, have cited Sedona Conference materials when resolving ediscovery disputes. Adopting the best practices identified in these materials can help increase the defensibility of your ediscovery methods.

5. DOCUMENT EVERYTHING AND BE PREPARED TO EXPLAIN YOUR PROCESS TO THE COURT.

Regardless of what steps you take to ensure the defensibility of decisions about technology use, it is important to document those steps and ensure that someone is able to explain them to the court. *Victor Stanley* emphasized that litigants also need to be able to demonstrate *why* their method was appropriate for the task and that it was properly implemented.

In this regard, expert testimony or assistance can be invaluable. In fact, at least one judge has repeatedly suggested that expert testimony is essential when the issue before the court is the appropriateness of a given search methodology. In a dispute over the proper search terms to be employed, Magistrate Judge John Facciola wrote:

> 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 key 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.

United States v. O'Keefe, 537 F. Supp. 2d 14 (D.D.C. 2008). *See also Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331 (D.D.C. 2008) (citing *O'Keefe* and requiring plaintiff to submit an affidavit from its forensic computer examiner before court would rule on a discovery dispute); *Eurand, Inc. v. Mylan Pharms., Inc.*, 266 F.R.D. 79 (D. Del. 2010) ("Neither lawyers nor judges are generally qualified to opine that certain search terms or files are more or less likely to produce information than those keywords or data actually used or reviewed.").

6. ASK THE COURT TO SIGN OFF ON A CLAWBACK AGREEMENT.

Federal Rule of Evidence 502(d) allows the parties to enter, and the court to sign off on, a "clawback agreement" providing that mistakenly producing documents will not effect a waiver of privilege. Such an agreement can ensure that inadvertently disclosed documents will retain privilege and will not be admissible in the current case or future cases.

Courts have encouraged the use of clawback agreements (see, e.g., *Doe v. Nebraska*, Nos. 4:09-cv-3266, 4:10-cv-3005, 8:09-cv-456, 2011 WL 1480483 (D. Neb. Apr. 19, 2011)), and a court can issue a court-ordered clawback agreement even if one party requests such an agreement and the other party opposes it. *See, e.g., Rajala v. McGuire Woods, LLP*, 2010 WL 2949582 (D. Kan. 2010).

By getting the court's approval for a clawback agreement, litigants can both preserve privilege and avoid having to prove that they took reasonable steps to avoid inadvertent disclosure, potentially saving considerable time and cost.

Conclusion

With litigants facing huge volumes of ESI and short discovery deadlines, using technology to search, review, and analyze ESI is now a critical discovery skill. To be successful, lawyers must not only keep abreast of the expanding range of search and review technologies, but understand how to use them in a defensible manner.

Though every discovery process is different, being mindful of the guidelines discussed above can help litigants ensure that their use of technology is not only cost-effective, but defensible.

1. Blair & Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System*, 28 Com. A.C.M. 289 (1985) (publication of the Association for Computing Machinery).

2. Discussing the advantages and disadvantages in of these technologies in detail is beyond the scope of this paper. For an in-depth discussion of the challenges inherent in using search and retrieval technologies in ediscovery, refer to *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-discovery*, available at http://www.thesedonaconference.org/content/miscFiles/publications_html.

3. *See, e.g., Kipperman v. Onex Corp.*, 2009 WL 1473708, Civ. Act. No. 1:05-CV-1242-JOF (N.D. Ga. May 27, 2009) (ordering $1,022,700 in monetary sanctions for discovery abuses); *Green v. Blitz U.S.A., Inc.*, 2011 WL 806011, No. 2:07-CV-372 (TJW) (E.D. Tex. Mar. 1, 2011) (ordering defendant to pay plaintiff $250,000 and provide a copy of the court's order to plaintiffs "in every lawsuit proceeding against it" for the past two years and to file the court's order in every case that defendant is involved in for the next five years).

4. *See, e.g., Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2008) (holding that defendants waived attorney-client privilege and work-product protection when they accidentally produced privileged documents because they failed to demonstrate that their ediscovery search methodology was defensible). In a subsequent opinion in the same case, the judge awarded $1,049,850.04 in attorney's fees and costs as a sanction for discovery abuses. *See Victor Stanley, Inc. v. Creative Pipe, Inc.*, No. MJG-06-2662 (D. Md. Jan. 24, 2011).

5. *See* Gibson and Dunn's 2011 Mid-Year E-Discovery Update (July 22, 2011) and 2010 Year-End Electronic Discovery and Information Law Update (Jan. 13, 2011), available at http://www.gibsondunn.com/publications/default.aspx.

6. *See, e.g., William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134 (S.D.N.Y. 2009) (selecting search terms to be used in response to a dispute about overly broad search terms); *Trusz v. UBS Realty Investors, LLC*, 3:09-CV-268, 2011 WL 1628005 (D. Conn. Apr. 27, 2011) (dictating search terms and date ranges to be used in response to a dispute about overly broad search terms).

7. This paper does not address technology assisted review (TAR). A discussion of the defensible use of TAR is beyond the scope of this paper.

8. One exception to this is *Disability Rights Council of Greater Wash. v. Wash. Metropolitan Transit Auth.*, 242 F.R.D. 139 (D.D.C. 2007), which Magistrate Judge John M. Facciola drew attention to concept searching as an alternative to keyword searching: "I bring to the parties' attention recent scholarship that argues that concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results."

9. 250 F.R.D. 251 (D. Md. 2008).

10. 244 F.R.D. 650 (M.D. Fla. 2007).

11. 256 F.R.D. 134 (S.D.N.Y. 2009).

12. Available at http://www.thesedonaconference.org/content/tsc_cooperation_proclamation/proclamation.pdf.

13. See http://www.thesedonaconference.org/content/tsc_cooperation_proclamation.

14. See http://www.thesedonaconference.org/publications_html?grp=wgs110.

About Lighthouse

For 25 years, Lighthouse has provided innovative software and services to manage the increasingly complex landscape of enterprise data for compliance and legal teams. Lighthouse leads by developing proprietary technology that integrates with industry-leading third-party software, automating workflows, and creating an easy-to-use, end-to-end platform. Lighthouse also delivers unique proprietary applications and advisory services that are highly valuable for large, complex matters, and a new SaaS platform designed for in-house teams. Whether reacting to incidents like litigation or governmental investigations, or designing programs to proactively minimize the potential for future incidents, Lighthouse partners with multinational industry leaders, top global law firms, and the world's leading software provider as a channel partner.

Connect with us to see how Lighthouse can best support you.

(206) 223–9690 | lighthouseglobal.com | info@lighthouseglobal.com

LIGHTHOUSE