Client Alert

September 2018

Don't Be Afraid of the Known Unknown – Validating Your Search Process

On August 7, 2018, the Northern District of Illinois issued an opinion that provides practical guidance regarding the validation of discovery processes and productions in *City of Rockford v. Mallinckrodt ARD Inc.*[1] Drafted with a welcome touch of whimsy by Judge Iain Johnston, the opinion speaks to a fundamental question in document investigations: How do we know we are ***done***? This is followed by a second, equally important question: Are we comfortable with what we have missed?

Judge Johnston incorporated elusion test principles frequently associated with Technology Assisted Review (TAR) into a traditional keyword searching protocol negotiated by the litigants. While TAR and keywords are often considered two different paths to responsive document retrieval, his unification of these two search processes ensures this ruling will drive discussions between litigants who intend to follow either route.

In *City of Rockford,* the parties had cooperatively developed a protocol to govern the use of keywords to search electronically stored information (ESI) and locate relevant documents they expected to find. But they could not resolve whether and how to search for "other" responsive documents that might or might not actually exist and that would not be identified by the keywords. Judge Johnston, quoting Donald Rumsfeld, characterized the issues as a search for "known unknowns"[2] that might demonstrate whether discovery is really complete.

Document production during litigation invokes Federal Rule of Civil Procedure 26(g)'s obligation to certify that, based upon a **reasonable inquiry**, any disclosure is complete and correct. Discovery protocols are intended to streamline this reasonable inquiry by identifying categories of documents that are likely relevant to the claims or defenses in the action. This process generally results in setting aside a "null set"[3] of files thought to possess very little value to the case before review begins (or during review, if using TAR). But the number of relevant documents possibly within that null set, and whether they contain unique substantive information or merely information also present in documents already identified, is a "known unknown."

Although "no one could or should expect perfection from [the search] process,"[4] the court found that the Rule 26(g) reasonable inquiry obligation extends to active validation and quality assurance beyond merely reviewing results of keyword searches. If you intend to review less than all of the potentially responsive data, whether by use of keyword searches or TAR, that decision should "be backed up with

---

[1] __F.R.D.__; 2018 WL 3766673 (2018).

[2] "[A]s we know there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know ...." Donald Rumsfeld. United States Department of Defense News Briefing (Feb. 12, 2002), http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636.

[3] A "null set" is defined as: "The set of Documents that are not returned by a search process or that are identified as Not Relevant by a review process." Source: Maura R. Grossman and Gordon V. Cormack, EDRM page & *The Grossman-Cormack Glossary of Technology-Assisted Review, with Foreword by John M. Facciola, U.S. Magistrate Judge*, 2013 Fed. Cts. L. Rev. 7 (January 2013).

[4] *Federal Housing Finance Agency v. HSBC North America Holdings, Inc.*, 2014 WL 584300, at *2-3, 2014 U.S. Dist. LEXIS 19156, at *32 (S.D.N.Y. Feb. 14, 2014).

appropriate facts that justify this choice. [While] generally, no single factor will be determinative on its own … demonstrably valid statistics can be one of the factors used to justify this decision."[5]
Keyword searching has become a common method to identify potentially relevant documents, though it will inevitably involve the retrieval of false positives—documents that are not relevant even though they contain the keyword—and the failure to identify 100 percent of relevant documents. Though TAR may be more efficient or accurate in identifying relevant data in some instances, as Judge Johnston points out, "the parties in this case have agreed to use keyword searching at the outset …. Courts cannot preach party cooperation and then unilaterally reject the reasonable process the parties agreed to use."

The court then promoted the benefits of an elusion test, a statistical analysis commonly used in TAR where review of a random sample from the null set determines how many (or what proportion of) documents presumed not to be relevant actually are.[6] The fraction of relevant documents found within the sample is reported as the elusion rate.

Based on the evidence presented by the parties in this case, the court ruled application of an elusion test[7] was both reasonable and proportional to identify the likelihood that responsive documents were "missed" by the agreed keywords. Any responsive documents discovered in the elusion test sample would be produced to the requesting party, who would be free to seek further discovery on that basis.

Interestingly, this court needed to find that the sampling both served as a reasonable inquiry **and** that it was proportional to the needs of the case. The court might not have ordered the elusion test if evidence had been presented to demonstrate an undue burden to the reviewing party conducting the elusion test. But in light of the evidence before it, the court found that because of the magnitude of the case, and in consideration of the other proportionality factors under Federal Rule of Civil Procedure 26(b)(1), a claim of undue burden or expense was not likely to carry the day in this particular case.

The court was clear that the burden of demonstrating reasonable inquiry rests squarely on the producing party. The opinion holds that one way the producing party can meet this burden is by crafting searches carefully and showing the validity of its underlying assumptions that all critically responsive documents have been located through the use of randomly sampling the documents left behind. This opinion promotes the proposition that design and validation of a reasonable process, rather than perfection during search, is what matters and is the basis for certifying compliance with Rule 26(g).

It is also important to note that while an elusion rate is designed to tell you what percentage of the null set documents are actually relevant, it does not tell either party whether those documents are critical enough to merit further discovery. The rate needs to be considered in the context of the responsive rate of documents already reviewed, the overall volume of the document population and the substantiveness of any responsive documents identified through the elusion sample. In some instances, a low elusion rate is a sign that the identification of relevant documents was sufficient, but in other cases the same elusion rate may indicate a need to adjust the search mechanism if the universe of documents is not rich with responsive documents. A skilled eDiscovery lawyer can help interpret the elusion rate, in conjunction with other metrics associated with processing and review of documents.

Authors

Corey Lee
leec@HuntonAK.com

Meghan A. Podolny
mpodolny@HuntonAK.com

[5] *EDRM Statistical Sampling Applied to Electronic Discovery*, Section 1, https://www.edrm.net/resources/project-guides/edrm-statistical-sampling-applied-to-electronic-discovery/.
[6] Maura R. Grossman and Gordon V. Cormack, EDRM page & *The Grossman-Cormack Glossary of Technology-Assisted Review, with Foreword by John M. Facciola*, U.S. Magistrate Judge, 2013 Fed. Cts. L. Rev. 7 (January 2013).
[7] The court ordered use of a statistically random sample size based on calculating a confidence level of 95 percent and a margin of error of 2 percent.

Michele R. Blythe
micheleblythe@HuntonAK.com

Joshua P. Hanbury
jhanbury@HuntonAK.com

Amanda Strassner Merrill
amerrill@HuntonAK.com