New York Law Journal
SERVING THE BENCH AND BAR SINCE 1888
WWW.NYLJ.COM

An incisivemedia publication
VOLUME 241—NO. 80
TUESDAY, APRIL 28, 2009

FEDERAL E-DISCOVERY

Expert Analysis

Court Issues 'Wake-Up Call' On Slipshod Search Terms


By **H. Christopher Boehning**
And **Daniel J. Toal**

Given the ubiquitous use of general-purpose search engines such as Google and attorneys' routine use of legal search engines such as Westlaw and Lexis-Nexis, it is perhaps surprising that lawyers frequently falter in formulating search terms, or "keywords," designed to retrieve relevant e-mails and other electronically stored information (ESI).

Nevertheless, courts have time and again confronted haphazard and uncoordinated search methodologies for ESI.

Evidently weary of deficient keyword searches, U.S. Magistrate Judge Andrew J. Peck recently issued a self-styled "wake-up call" to members of the bar in the Southern District. Instead of attorneys designing keywords without adequate information "by the seat of their pants," Magistrate Judge Peck appealed for keyword formulations based on careful thought, quality control, testing and cooperation.

The magistrate judge's admonition arose in *William A. Gross Constr. Assocs., Inc. v. American Mfrs. Mut. Ins. Co.*[1] The case involved multiple parties and multi-million dollar claims concerning alleged defects and delays in the construction of the Bronx County Hall of Justice.

The Dormitory Authority of the State of New York (DASNY), a public benefit corporation that acts as the developer of courthouses, directed the project. At the time of the discovery dispute, non-party Hill International served as DAS-NY's construction manager. DASNY consented to produce Hill's project-related documents and ESI to the other parties in the action.

> Lawyers well acquainted with computers may be more susceptible to thinking that keywords viable for a Google search should also suffice for ESI production. Southern District Magistrate Judge Andrew J. Peck has sounded the alarm that such haphazard searches will not pass muster any longer.

As obviously relevant ESI, Hill's e-mails presented a classic challenge of devising a proper search methodology for production. Hill understandably did not want to produce e-mails unrelated to the Bronx courthouse project, but combing through the e-mails one by one to cull unrelated e-mails would have been time-consuming and uneconomical.

DASNY proposed the following search terms to collect the relevant e-mails: "DASNY," "Dormitory Authority," "Authority" and the names of the other parties in the action. In addition, DASNY suggested "Court! in connection with Bronx," "Hall of Justice" and "Bronx but not Zoo"—to distinguish e-mails relating to Hill's work on a Bronx Zoo project. The other parties sought a litany of additional search terms, running into the thousands. Their terms corresponded to the construction issues involved in the courthouse project, such as "sidewalk," "delay," budget," "elevator," "claim" and the like, which, when applied to a construction management business such as Hill, threatened to require production or manual review of Hill's entire e-mail database.

Despite Hill's unique ability to explain the argot used in its employees' e-mails to the court and the parties, Hill proved relatively unhelpful in resolving the debate. On the one hand, Hill agreed with the other parties that DASNY's search terms were likely too narrow; on the other hand, Hill considered the other parties' terms too broad. But Magistrate Judge Peck found that Hill otherwise offered no assistance in developing appropriate search terms.

The magistrate judge was therefore placed in the "uncomfortable position"

H. CHRISTOPHER BOEHNING *and* DANIEL J. TOAL *are litigation partners at Paul, Weiss, Rifkind, Wharton & Garrison LLP.* YITZCHAK E. SOLOVEICHIK, *a law clerk at the firm, assisted in the preparation of this article.*

of having to construct a search term methodology without sufficient input from the parties or the relevant custodian. Accordingly, he ruled that in addition to DASNY's proposed terms, the search should incorporate the names of the parties' personnel involved in the courthouse project. Magistrate Judge Peck conceded in a footnote that this result was less than perfect and might require modification based on the results of discovery.

Describing this case as "just the latest example of lawyers designing keyword searches in the dark," without adequate discussion with those who wrote the e-mails, Magistrate Judge Peck took the opportunity to reiterate prior warnings about this problem from judges in the Baltimore-Washington Beltway. In his view, these prior warnings had not gotten through to the bar in the Southern District. The earlier warnings were tailored to the different circumstances of those cases, but Magistrate Judge Peck apparently thought them equally applicable across the spectrum of electronic discovery issues.

As Magistrate Judge Peck noted, one such warning was issued in *Victor Stanley v. Creative Pipe, Inc.*[2]

In *Victor Stanley*, the defendants inadvertently produced attorney-client privileged ESI fles and sought "clawback" approval from the court, arguing that privilege had not been waived because the disclosure was inadvertent.[3] Magistrate Judge Paul W. Grimm ruled that the inadvertent production of privileged materials waived the privilege because the defendants did not demonstrate that they had taken reasonable precautions to prevent inadvertent disclosure.

Because the limitations and risks associated with keyword search methods for ESI require "technical, if not scientifc knowledge" to achieve proper selection and implementation of the keyword search, Magistrate Judge Grimm criticized the defendants' failure to explain the keywords used, why they were chosen, the qualifcation of the keyword creators, and whether the methodology had been tested for reliability. The court emphasized that individuals qualified to design search criteria must engage in careful planning before finalizing a search method. In addition, he cautioned that the party selecting the search technique must be capable of explaining the rationale behind it, demonstrating its appropriateness and proving its proper implementation.

More Than Guesses

In *William A. Gross*, Magistrate Judge Peck endorsed Magistrate Judge Grimm's description of the proper procedure for devising search keywords. He emphasized in a footnote that what is required is more than a lawyer's guesses, without any quality control testing to ensure the search results are minimally over- and under-inclusive for responsive e-mails.

Accordingly, care should be taken from the outset of disclosure to ensure proper selection and execution of a search methodology. Without appropriate care, a court cannot be confident that the producing party has disclosed all the required responsive material.

Another note of warning was sounded by Magistrate Judge John M. Facciola in *United States v. O'Keefe*.[4]

In O'Keefe, the indictment charged one defendant with having received gifts from his co-defendant in exchange for expediting visa requests while an employee of the Department of State in Canada. An earlier court order had required the government to make a good faith effort to uncover all responsive information in the hard copy and electronic files of various consulates. Upon the government's submission of documents in compliance with this order, the defendants moved to compel, claiming that the government had not discharged the order's obligations.

With regard to the government's electronic production, the defendants advanced several objections, including against the government's keyword methodology. In his ruling, Magistrate Judge Facciola phrased this objection in two subtly different ways. Initially, when listing the defendants' various objections, he described their keywords objection as protesting that the government did not indicate "how it ascertained what search terms it would use."[5] When turning to a detailed discussion of this particular objection, he stated: "As noted above, defendants protest the search terms the government used."[6] The significance of this difference in formulation becomes clear in light of Magistrate Judge Facciola's ruling.

Magistrate Judge Facciola deemed the issue of whether search terms will yield the information sought to be a complicated question involving the interplay, "at least," of computer science, statistics, and linguistics. He continued, in a passage quoted by Magistrate Judge Peck in *William A. Gross*:

> 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.[7]

Magistrate Judge Facciola ruled that any claim by the defendants that the government's search terms were insufficient would need to specifcally so contend based on evidence that

meets the requirements of Rule 702.

This ruling highlights the significance of Magistrate Judge Facciola's two formulations of the defendants' objection. The frst formulation—that the government failed to indicate how it ascertained the search terms used—does not object to the particular search terms per se. The formulation allows for the possibility that the search terms are suffcient, but the objector seeks some rationale for their selection. By contrast, the second formulation—the "defendants protest the search terms the government used"—strongly implies qualms with the search terms in and of themselves.

But Magistrate Judge Facciola used the two formulations interchangeably. He did so, it seems, because in his view there is no separate sustainable objection that the producing party has failed to explain its choice of search terms. To object to search terms at all, he requires that the objection be to the search terms per se. The objecting party must explain in detail—and possibly with Rule 702 evidence, no less—why the search terms are inadequate.

In other words, the ruling in *O'Keefe* places the not insignificant burden of investigating the character of the ESI and meticulously formulating a proper keyword methodology for the ESI on the objecting party rather than on the producing party.

Magistrate Judge Facciola reinforced this facet of his *O'Keefe* ruling less than a month later. In *Equity Analytics, LLC v. Lundin*,[8] the plaintiff contended that nokeyword search methodology would be adequate to gather the relevant ESI files. Citing *O'Keefe*, Magistrate Judge Facciola required the plaintiff to submit an affidavit from an expert explaining why the defendant's proposed search methodology would be inadequate. Once again, he placed the burden of explaining the proper search methodology on the objecting party.

Magistrate Judge Peck's "wake-up call" in *William A. Gross* differs from the warnings issued by Magistrate Judge Facciola in two important ways. First, Magistrate Judge Peck appears less inclined to require expert testimony. In a footnote, he observed that he did not need to decide at that time whether expert testimony was required.

Second, in Magistrate Judge Peck's view, the producing party bears the burden of formulating a reasoned search methodology. As he quoted approvingly from *Victor Stanley*, "the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court."[9] His quotation of *O'Keefe* and citation to *Equity Analytics* do not appear designed to suggest otherwise. Both cases appear to have been referenced because they explain the difficulties posed by keyword searches of ESI, and for that reason alone.

Nonetheless, given his lengthy quotation of *O'Keefe*, Magistrate Judge Peck's reluctance to remark upon the significant apparent difference of opinion between himself and Magistrate Judge Facciola is rather surprising. A matter of such import—that is, which party bears the burden of offering a reasoned search methodology—seems worthy of explicit discussion.

Conclusion

Magistrate Judge Peck's opinion in *William A. Gross* stressed four requirements for the production of ESI. Foremost, there must be cooperation between opposing counsel. Therefore, he strongly endorsed The Sedona Conference Cooperation Proclamation.

Second, attorneys must carefully design the appropriate keywords. Third, these keywords should be selected with the input from the ESI's custodians. Finally, the proposed technique should be validated to ensure it is not substantially over- or under-inclusive.

Magistrate Judge Peck concluded with the following admonition: "It is time that the Bar—even those lawyers who did not come of age in the computer era—understand "the importance of properly crafted electronic searches.

Ironically, lawyers well-acquainted with computers may be more susceptible to thinking that keywords viable for a Google search should also suffice for ESI production. Magistrate Judge Peck has sounded the alarm that such haphazard searches will not pass muster any longer.

1. No. 07 Civ. 10639, 2009 WL 724954(S.D.N.Y. March 19, 2009).
2. 250 F.R.D. 251 (D. Md. 2007).
3. For a more detailed discussion of Victor Stanley, see H. Christopher Boehning and Daniel J. Toal, "Poorly Executed Privilege Review Can Lead to Waiver," NYLJ, June 17, 2008.
4. 537 F. Supp. 2d 14 (D.D.C. 2008).
5. Id. at 22.
6. Id. at 23.
7. Id. at 24.
8. 248 F.R.D. 331 (D.D.C. 2008).
9. *William A. Gross*, 2009 WL 724954, at *2.