

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

USDC SDNY
DOCUMENT
ELECTRONICALLY FILED
DOC#
DATE FILED: 3/19/09

----- x

WILLIAM A. GROSS CONSTRUCTION :
ASSOCIATES, INC., :

Plaintiff, :

-against- :

AMERICAN MANUFACTURERS MUTUAL :
INSURANCE COMPANY, :

Defendant. :

----- x

07 Civ. 10639 (LAK) (AJP)

OPINION AND ORDER

ANDREW J. PECK, United States Magistrate Judge:

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 or "keywords" to be used to produce emails or other electronically stored information ("ESI"). While this message has appeared in several cases from outside this Circuit, it appears that the message has not reached many members of our Bar.

FACTS

This case involves a multi-million dollar dispute over alleged defects and delay in the construction of the Bronx County Hall of Justice, also known as the Bronx Criminal Court Complex. See William A. Gross Constr. Assocs., Inc. v. American Mfrs. Mut. Ins. Co., 07 Civ. 10639, 2009 WL 427280 (S.D.N.Y. Feb. 23, 2009) (Peck, M.J.). The Dormitory Authority of the State of New York ("DASNY") was the "owner" of the project. Non-party Hill International is DASNY's current

CAOPINGROSS-EDisc

construction manager, and DASNY has agreed to produce Hill's project-related documents and ESI to the other parties to this suit. The issue before the Court is the production of Hill's emails, stored on its server at its New Jersey office, and how to separate project-related emails from Hill's unrelated emails.

DASNY's proposed search terms included "DASNY," "Dormitory Authority," and "Authority,"^{1/} and the names of the other parties to the action. DASNY also added "Court! in connection with Bronx," "Hall of Justice," and "Bronx but not Zoo" (since Hill worked on an unrelated project involving the Bronx Zoo). The other parties requested the use of thousands of additional search terms, emphasizing the construction issues they were involved in, such as "sidewalk," "change order," "driveway," "access," "alarm," "budget," "build," "claim," "delay," "elevator," "electrical" – you get the picture. DASNY correctly pointed out that use of such extensive keywords would require production of the entire Hill email database, since Hill's business is construction management, and those terms would be used for any construction project.

Hill's only contribution to the discussion was to agree that DASNY's search terms were probably too narrow but the other parties' terms were overbroad, and that Hill did not want to produce emails that did not relate to the Bronx Courthouse project. This problem would have been avoided, of course, if Hill used a standard "Re" line in its Bronx Courthouse emails to distinguish that project from its other work. It did not do so, however. Moreover, while Hill was in the best position to explain to the parties and the Court what nomenclature its employees used in emails, Hill

^{1/} The Court is no keyword expert, but if one is searching for "Authority," to also search for "Dormitory Authority" is clearly redundant.

did not do so – perhaps because, as a non-party, it wanted to have as little involvement in the case as possible.

That left the Court in the uncomfortable position of having to craft a keyword search methodology for the parties, without adequate information from the parties (and Hill). The Court ruled at yesterday's conference that in addition to DASNY's proposed terms (including variations on and abbreviations of party names), the search should also include the names of the parties' personnel involved in the Bronx Courthouse construction.^{2/}

DISCUSSION

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. Prior decisions from Magistrate Judges in the Baltimore-Washington Beltway have warned counsel of this problem, but the message has not gotten through to the Bar in this District. As Magistrate Judge Paul Grimm has stated:

While keyword searches have long been recognized as appropriate and helpful for ESI search and retrieval, there are well-know limitations and risks associated with them, and proper selection and implementation obviously involves technical, if not scientific knowledge.

* * *

Selection of the appropriate search and information retrieval technique requires careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected should be tested for

^{2/} The Court acknowledges that this result is less than perfect, and that there is a risk that as information later comes out at depositions of the Hill employees, another search may have to be done.

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.

Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260, 262 (D. Md. May 29, 2008) (Grimm, M.J.).

Magistrate Judge Facciola has taken the warning even further:

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 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.

United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (Facciola, M.J.); accord, Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008) (Facciola, M.J.);^{3/} see also, e.g., In re Seroquel Products Liability Litig., 244 F.R.D. 650, 662 (M.D. Fla. 2007) (Baker, M.J.) ("[W]hile key word searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process. . . . Common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness."); Jay Grenig, Browning Marean & Mary Pat Poteet, Electronic Discovery & Records Management Guide: Rules, Checklists & Forms (2009 ed.), § 15:15 ("[K]eyword searches do not

^{3/} This Court need not now decide whether expert testimony is required; what is required is something other than a lawyer's guesses, without client input, and without any quality control testing to see if the search terms produce reasonably all the responsive ESI and limited "false positives."

reflect context. They can also miss documents containing a word that has the same meaning as the term used in the query but is not specified. Misspelled words may be missed in a keyword search.").

Of course, the best solution in the entire area of electronic discovery is cooperation among counsel. This Court strongly endorses The Sedona Conference Cooperation Proclamation (available at www.TheSedonaConference.org).

CONCLUSION

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.

SO ORDERED.

Dated: New York, New York
March 19, 2009



Andrew J. Peck
United States Magistrate Judge

Copies **by ECF** to: All Counsel
Judge Lewis A. Kaplan