Proskauer

Court Endorses Use of Predictive Coding for ESI Discovery

March 9, 2012

In a recent opinion *Da Silva Moore v. Publicis Groupe & MSL Group*, 11-CV-1279 (S.D.N.Y. Feb. 25, 2012), Magistrate Judge Peck recognized the use of predictive coding technology, also referred to as computer-assisted review, as an appropriate method to satisfy a producing party's review obligations in appropriate cases.

Factual Background

In the action, five female named plaintiffs are suing Publicis Groupe and its United States public relations subsidiary, MSL Group (collectively, "Defendants"). Plaintiffs allege that defendants have a "glass ceiling" that limits women to entry-level positions, and that there is systemic, company-wide discrimination against female public relations employees.

In this case, the parties were faced with the daunting task of reviewing approximately three million documents from the agreed-upon custodians.

Computer-Assisted Review Explained

Computer-Assisted Review refers to tools that use sophisticated algorithms to enable the computer to determine relevance of a document, based on interaction with a human reviewer. It involves manually reviewing and coding a "seed set" of documents[1], thereby enabling the computer to identify properties of those documents that it then uses to code other documents. As the reviewer continues to code more sample documents, the computer predicts the reviewer's coding.

When the system's predictions and reviewer's coding become sufficiently consistent, the system has learned enough to make confident predictions as to the remaining documents.[2]

Some systems produce a simple yes/no as to relevance, while others give a relevance score (e.g., on a 0 to 100 basis) that can be used to prioritize review. The example given by the Court was that a score above 50 may produce 97% of the relevant documents, but constitute only 20% of the entire document set. Additionally, counsel may decide that documents below a certain score are so unlikely to be relevant that no human review is necessary, saving significant time and costs to an employer.

The Southern District's Ruling

"What the Bar should take away from this Opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review." *Id*. at 25.

The Southern District found computer-assisted review, in appropriate cases, to be a preferable option to most if not all of the available alternatives, and at a fraction of the cost. *Da Silva Moore* was thought to be an appropriate case as it would require the review of millions of documents. The Court noted that "Courts and litigants must be cognizant of the aim of Rule 1, to 'secure the just, speedy, and inexpensive determination' of lawsuits.'" *Id*. at 22. The Court insists this is not a matter of machines replacing humans, but more an interaction of man and machine, one that can prove extremely useful in more efficiently and cost-effectively reaching the parties' goals.

Judge Peck determined that the use of predictive coding technology was appropriate in this case considering the following five factors:

"(1) the parties' agreement,

(2) the vast amount of ESI to be reviewed (over three million documents),

(3) the superiority of computer-assisted review to the available alternatives (i.e., linear manual review or keyword searches),

(4) the need for cost effectiveness and proportionality under Rule 26(b)(2)(C), and

(5) the transparent process proposed by [the defendant]."

(*Id*. at 22).

As noted above, in *Da Silva Moore* the parties agreed to the computer-assisted review, although they disagreed about the manner in which it should be implemented. The Southern District anticipated that all cases would not involve such a mutual agreement. Judge Peck found that in contested situations the question the Court should ask is, what methodology would the requesting party suggest instead? Manual review of millions of emails is simply too expensive a task. Additionally, the Court cited evidence to refute the "myth" some espouse, which is that manual review is more accurate or somehow preferable to computer-assisted review. In fact, the Court noted in its opinion that statistics clearly show that computerized searches are at least as accurate, if not more so, than manual review. So, while the parties' agreement was a factor which influenced the Court's decision, Judge Peck's opinion suggests that a request for computer-assisted review would be appropriately granted in certain large data volume cases even where one party objected to its use. This is a significant step made by the Southern District in recognizing the immense cost of certain types of litigation to employers.

Implications for Employers

Clearly, this decision is significant to any employer who has in the past, or may in the future, be involved in a lawsuit involving large amounts of ESI. Most often this type of situation would arise in the context of a putative class or collective action. In such cases, there is a great potential for cost-saving by using predictive coding technology. Employers would be wise to become familiar with the technology and assess whether or not it could be of use in their current or future legal matters.

[1] In this case, the parties agreed to use a 95% confidence level to create a random sample of the entire email collection; that sample of 2,399 documents will be reviewed to determine relevant documents for a "seed set" to use to train the predictive coding software.

[2] Typically, a few thousand documents must be reviewed to train the computer.

Related Professionals

- **Steven D. Hurd**
  Partner
- **Mark W. Batten**

Partner