

New Jersey Law Journal

VOL. 211 - NO 1

JANUARY 7, 2013

ESTABLISHED 1878

IN PRACTICE

E-DISCOVERY

Predictive Coding Technology: The New Frontier

By A. Ross Pearlson and
James Van Splinter

Predictive coding technology — the use of computer software to assist and expedite discovery document review — is the latest e-discovery tool with which litigation counsel must become familiar. By potentially eliminating the need for large-scale document review, predictive coding technology can reduce costs and streamline the e-discovery process. Proponents also argue that computer-assisted review can be more accurate than traditional attorney review. However, as with any new technology, counsel must be prepared to assess its usefulness in a particular case and to address potential client concerns.

Parties implementing predictive coding: (i) start with a set of data, derived from “traditional” methods (e.g., keyword searching); (ii) review a “seed” or sample set of documents for responsiveness and/or privilege; (iii) employ computer-learning software to review and categorize potentially responsive documents; and (iv) analyze the results to obtain quality-control feedback and achieve coding consistency. This process

Pearlson is co-chairman of the litigation group at Wolff & Samson PC in West Orange. Van Splinter is an associate in the group.

is repeated until the party is confident that the results of the computer-assisted review gibe with the responsiveness and privilege determinations made through counsel’s review of the seed set.

Predictive coding in e-discovery was recently addressed by Magistrate Judge Andrew Peck in *Da Silva Moore v. Publicis Groupe*, No. 1:11-cv-01279 (S.D.N.Y. Feb. 24, 2012), in the context of a gender discrimination suit against France-based Publicis Groupe and its U.S. subsidiary MSL Group. Judge Peck, coincidentally, had previously voiced support for the use of predictive coding technology. See Hon. Andrew Peck, “Search, Forward,” *L. Tech. News* at 29 (Oct. 2011) (stating that “computer-assisted coding should be used in those cases where it will help secure the just, speedy, and inexpensive ... determination of cases.”).

The parties raised the potential use of predictive coding technology at their first conference before Judge Peck and hired technology consultants to help prepare a predictive coding protocol. The parties jointly submitted a 22-page protocol, which identified 28 sources of electronically stored information (ESI) (such as emails, computers and removable storage devices) and 31 document custodians. The proposed protocol identified the software to be used by MSL and the procedure for MSL’s counsel’s

initial review and categorization of the sample set of documents. Plaintiffs’ counsel would provide their own proposed keywords for the seed-set search, and MSL would provide nonprivileged documents identified both as “responsive” and “nonresponsive” following the seed-set review conducted by senior attorneys. Plaintiffs’ counsel would then advise whether the documents were properly coded, with any disputes to be resolved by the court. The protocol also included a confidentiality/clawback agreement.

The next step under the protocol was a “training” process in which a batch of documents would be computer-reviewed and subjected to attorney review for quality control (during which plaintiffs’ counsel could again provide input as to relevancy). The protocol included seven rounds of iterative software training and a 5 percent target error rate. If, after the seventh iteration, there was a greater than 5 percent change in documents coded as responsive as compared to the sixth iteration, the training process would not be complete. If the target error rate was below 5 percent, then the training process would be deemed complete, and MSL could apply the coding procedures to the entire data set.

The protocol provided a quality-control process whereby MSL would review a batch of documents coded by the software as nonresponsive and provide the results to the plaintiffs.

The plaintiffs objected to the protocol, arguing that: (1) it would not ensure “complete” production of documents

pursuant to F.R.C.P. 26(g)(1)(A); (2) the predictive coding technology did not comport with the requirements of F.R.E. 702; and (3) the reliability of MSL's proposed coding process was impossible to determine.

Judge Peck upheld the use of the protocol, noting that F.R.C.P. 26(g)(1)(A) applies only as to mandatory initial disclosures, while document discovery is governed by F.R.C.P. 26(b)(2)(C), which contains a "proportionality doctrine" balancing the need for discovery against the costs of production. Judge Peck rejected the plaintiffs' F.R.E. 702 arguments because that rule and *Daubert* concern the reliability of expert testimony, not the efficacy of e-discovery technology. Judge Peck also ruled that the plaintiffs' reliability concerns were premature because the "training" process, which was open to the plaintiffs' review and comments, had not yet begun.

Judge Peck's ruling was also based on: (1) the parties' agreement; (2) the large amount of ESI to be reviewed; (3) the superiority of computer-assisted review over alternatives such as manual review or keyword searches; (4) its compliance with F.R.C.P. 26(b)(2)(C)'s proportionality requirement; and (5) the transparency of the process.

On appeal, Judge Carter upheld Judge Peck's ruling. Notably, Judge Carter ruled that the plaintiffs' refusal to consent to the protocol "is immaterial because the ESI protocol contains standards for measuring the reliability of the process and the protocol builds in levels of participation by Plaintiffs." See *Da Silva Moore*, 1:11-cv-01279, at 3. (S.D.N.Y. April 26, 2012). Judge Carter also noted that "the parties are allowed to reconsider their methods and raise their concerns with the Magistrate Judge" if the coding software proves to be flawed or if plaintiffs feel they are not receiving the types of documents that should be produced.

The five-factor test articulated by

Judge Peck and affirmed by Judge Carter in *Da Silva Moore* addresses party consent, F.R.C.P. 26(b)(2)(C), and concerns regarding the scope, reliability and cost of using predictive coding technology as a method of document review. The test therefore appears to "cover all of the bases" and will, with further case law development, represent a useful standard for future litigants. Judge Peck's status as an authority in e-discovery provides the test with additional credence. With this in mind, litigants and counsel should consider the following issues:

1. *Consent*: Although consent is the first factor of the *Da Silva Moore* test, Judge Carter ultimately found the plaintiffs' consent irrelevant because the protocol was fundamentally fair and comported with the federal rules. In one of the few other cases to address predictive coding, a Virginia court ordered the use of predictive coding over the plaintiffs' unequivocal objections. See *Global Aerospace v. Landow Aviation*, Case No. CL 614040, (Va. Cir. Court, Loudoun Cty. 2012). Although it is difficult to draw conclusions from two decisions, it seems that where a court concludes that a protocol is otherwise acceptable, a party's lack of consent will not prevent its implementation. Thus, counsel must be prepared to address any proposed predictive coding protocol on its merits.

2. *Volume of Documents*: In *Da Silva Moore*, the predictive coding protocol included a seven-step iterative training process, senior attorney review of thousands of documents and the coding of over 3 million documents. Such an extensive and expensive process is not appropriate for smaller productions, which could not provide the software with a sufficient base set to "learn" the coding provided by counsel.

3. *Superiority of Computer-Assisted Review*: Judge Peck did not specify why MSL's proposed protocol (as opposed to the general concept of computer assisted review) was superior to more traditional

review processes, such as a managed review or keyword searching. Instead, Judge Peck ruled that the plaintiffs' "reliability" arguments could be addressed after the process had been completed. However, counsel should be prepared to defend the reliability of their protocol to a potentially more skeptical judge from the outset. Thus, although Judge Peck was correct that Rule 702 and *Daubert* were not applicable, it may be reasonable to conclude that a "*Daubert*-like" test concerning the reliability of a proposed review protocol (and the court's gatekeeper role) will be developed in the future.

4. *Cost Effectiveness and Proportionality*: Judge Peck determined that, given the large number of documents to be reviewed and the "superiority" of predictive coding review as compared to more traditional search methods, the requirements of F.R.C.P. 26(b)(2)(C) were met. As with factor 3 above, litigants should be prepared to demonstrate and defend the cost-effectiveness of their protocol.

5. *Transparency*: The transparency of MSL's proposed protocol was a major factor in favor of its implementation — the plaintiffs avoided any potential prejudice by being involved in the responsiveness and relevancy determinations such that their input would lead to the ultimate accuracy of the process. However, the transparency of the process and adverse counsel's participation in it are likely to raise the concerns of more than one client. In cases involving multiple document repositories and multiple custodians, many sensitive, nonresponsive documents are interspersed with those that are responsive. Any party wishing (or compelled) to engage in a "transparent" predictive search process must first consider and address potential issues implicated by an adverse party's review of confidential documents, and take care to establish a confidentiality agreement that fully addresses these concerns. ■