# Client Alert

Litigation Group                                                                 March 2012

## Federal Court Endorses
## Computer-Assisted Review in E-Discovery

Clients facing litigation with large amounts of electronic discovery need a new approach to reduce the legal costs of traditional human-based linear review. Results of several studies suggest that computer-assisted review could offer significant cost and time savings with less human error than linear review.[1] Despite the promising results from these studies, we were concerned about the efficacy of computer-assisted review and whether we could defend it in court if challenged after the fact.[2] On February 24, 2012, Magistrate Judge Peck from the United States District Court, Southern District of New York, became the first judge to expressly endorse the parties' use of computer-assisted document review.[3] The decision promises to increase the use of technology-assisted review and makes it more defensible.

The plaintiffs in *Da Silva Moore v. Publicis Groupe & MSL Group*, 11 Civ. 1279 (ALC)(AJP) (S.D.N.Y. Feb. 24, 2012) brought a class action discrimination case under Title VII and the Family and Medical Leave Act. In response to the plaintiffs' initial document demands, the defendant asserted that it had approximately three million documents that it needed to review. As a result, the defendant requested that it be allowed to use a computer-assisted review method known as "predictive coding."

Under the predictive coding method, senior attorneys would review and manually tag a "seed set" of 2,399 randomly selected documents. The seed set would be analyzed by a computer to identify properties of those documents that would be used to code other documents. Next the computer would use those patterns to produce about 500 "judgment based sample" documents, which would also be reviewed by the senior

---

[1] For discussion of these studies, *see* Chapter 8 on "Responding to Request for Disclosure of Electronically Stored Information" in Bisceglie, Kyle NEW YORK E-DISCOVERY AND EVIDENCE (LexisNexis 2012 Ed.).

[2] Critics have raised questions about the data set used in several of these studies. Some were performed on the Enron Corpus (*i.e.*, a freely available database of over 600,000 emails acquired by the Federal Energy Regulatory Commission during its investigation of Enron after the company's collapse) or other data sets where much was already known about the data population when planning the review.

[3] Judge Peck had recently authored an article in the Legal Technology News regarding predictive coding. Andrew Peck, *Search Forward: Will manual document review and keyword searches be replaced by computer-assisted coding?*, L. Tech. News (Oct. 2011). Judge Peck quoted extensively from this article along with an article from Maura Grossman & Gordon Cormack. *See* Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review*, Rich. J. L. & Tech., Spring 2011.

attorneys. The judgment based sample documents would be re-processed by the computer to determine where the computer and the human coders disagree. The computer would then produce another set of 500 documents and the process would be repeated either until six additional iterations have occurred or the variance between the human coders and the computer was below 5%, whichever came first. By following this method, attorneys would calibrate the algorithm the computer would use to identify relevant documents. Once the algorithm was satisfactorily refined, the computer would review all documents rapidly to identify the relevant documents.

In *Da Silva Moore*, both parties agreed that the defendant should utilize some form of predictive coding, but they disagreed on the details. The court agreed that this was a good case for predictive coding, and stepped in to help resolve the parties' differences over the specifics of the methods to be employed. Consistent with the court's prior decision encouraging cooperation in discovery (*see William A. Gross Constr. Assocs., Inc. v. Am. Mfrs.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009), the court praised the defendant for being open about its methods and coding. The defendant agreed to turn over all non-privileged documents in the "seed set" and each of the "judgment based sample" sets providing both relevant and irrelevant documents. The defendant also incorporated any coding changes proposed by the plaintiffs based on their review of these documents. In the court's view, openness was essential when using predictive coding. As a guide for other litigants, the court annexed to the opinion the parties stipulated protocol for utilizing the predictive coding. Click here for a copy of the decision and exhibit.

The court noted that "computer-assisted review works better than most of the alternatives, if not all the [present] alternatives. So the idea is not to make this perfect, it's not going to be perfect. The idea is to make it significantly better than the alternatives without nearly as much cost." The court favorably noted studies showing that manual linear review is not only expensive and slow, but also not necessarily as accurate as a computer-assisted review. The court noted that linear reviews become less accurate due to human errors and disagreements between reviewers, and that keyword searches are often "the equivalent of the child's game of 'Go Fish'" where the parties are merely guessing at keywords that might produce evidence. The court also noted that the Federal Rules of Evidence do not require a party to certify that its production is complete or perfect, rather courts apply the Rule 26(b)(2)(C) proportionality principle, taking into account factors such as the reasonableness of the request, its potential benefit to the requestor and the burden or expense placed on the producing party.

In the end, the court found predictive coding appropriate given: (1) the parties' agreement to use it, (2) the vast amount of electronically stored information to be reviewed, (3) the superiority of computer-assisted review over the available alternatives such as linear manual review or keyword searches, (4) "the need for cost effectiveness and proportionality under rule 26(b)(2)(c)," and (5) the transparent and open process proposed by the defendant.

E-discovery software will continue to improve in its ability to identify relevant documents.  In addition to giving the judicial imperator if not encouragement to use computer-assisted review, the *Da Silva Moore* decision provides guidance for many of the factors that should be taken into account.  This decision is already having a potentially paradigm shifting impact.  Indeed, we already have received requests from the government on active cases to use predictive coding citing *Da Silva Moore*, despite the absence of any discussion of computer-assisted review in the "Recommendations for Electronically Stored Information (ESI) Discovery Production in Federal Criminal Cases," which was issued on February 13, 2012 after 18 months of negotiating and drafting by the Administrative Office/Department of Justice Joint Electronic Technology Working Group.

We continue to approach computer-assisted review with cautious optimism.  Computer-assisted review is still largely untested in real cases, a fact noted by the court in *Da Silva Moore*.  It will involve substantially higher vendor costs both in terms of loading and processing documents and managing and refining of the computer system.  As a result, its use on smaller cases may not be justified.  Most of all, the *Da Silva Moore* decision suggests it is prudent to obtain your adversary's input and court's consent before using computer-assisted review to make your production.

Please feel free to contact the attorneys listed below or any partner with whom you work if you wish to discuss the implications of computer-assisted review.

_____

Kyle C. Bisceglie                          Mason A. Barney
kbisceglie@olshanlaw.com          mbarney@olshanlaw.com
212.451.2207                               212.451.2267