

# The Metropolitan Corporate Counsel®

www.metrocorpcounsel.com

Volume 20, No. 6

© 2012 The Metropolitan Corporate Counsel, Inc.

June 2012

## Panel Discussion: Judge Peck, *Da Silva Moore* And The Outlook For Predictive Coding

On February 24, in the first decision of its kind, *Monique Da Silva Moore, et al. v. Publicis Groupe and MSL Group* (Opinion), Magistrate Judge Andrew Peck of the Southern District of New York explicitly approved the use of predictive coding in electronic discovery. Judge Peck stated that “computer-assisted review can now be considered judicially approved for use in appropriate cases.” He added: [Predictive coding] certainly works better than most of the alternatives, if not all of the 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 alternative without nearly as much cost.” Following an appeal by the plaintiffs, Judge Peck’s



**Conor R. Crowley**



**Stephen J. Goldstein**



**Sean P. Foley**

completely mobilized the industry. In this panel, we will be trying to understand why and what the impact will be.

Working with Equivio, we have assembled the following distinguished panelists: **Conor R.**

**Crowley**, Crowley Law Office; **Stephen J. Goldstein**, Squire, Sanders & Dempsey L.L.P.; and **Sean P. Foley**, ProSearch Strategies, Inc.

**Editor: Let’s start with basics. What exactly is predictive coding? Is there**

decision was confirmed on April 26 by U.S. District Court Judge Andrew Carter.

Judge Peck’s opinion has attracted unprecedented attention in the legal community. In fact, it’s hard to remember an event in the history of e-discovery that has generated so much excitement and so

**Conor R. Crowley, Esq.**, CIPP/US, the founder of Crowley Law Office, advises corporate and law firm clients on e-discovery, information governance and data privacy. He is the Vice-Chair of The Sedona Conference® Working Group on Electronic Document Retention and Production, Editor-in-Chief of The Sedona Conference Commentary on Proportionality in E-Discovery, and Senior Editor of a number of The Sedona Conference’s publications including The Sedona Conference Commentary on Legal Holds and The Sedona Principles (Second Edition). He is an inaugural member of both the Advisory Board for Georgetown University Law Center’s Advanced E-Discovery Institute and the Board of Advisors for BNA’s Digital Discovery & e-Evidence, and a member of The Sedona Conference Working Group on International Electronic Information Management, Discovery and Disclosure, and the International Association of Privacy Professionals.

**Stephen J. Goldstein** serves as the Director of Practice Support for Squire Sanders. Since joining the firm in September 2001, he has managed the litigation support function including large-scale projects, such as those related to the Enron Unsecured Creditors Committee. Acting as an internal consultant to the firm’s lawyers, he provides technical guidance and solutions to client-related matters around the globe. With more than 15 years’ experience working with legal technology, he is a frequent speaker and presenter on topics such as litigation technology, electronic data discovery and predictive coding/technology-assisted review. Mr. Goldstein is a contributing author to the Electronic Data Reference Model (EDRM) project and a member of the Sedona Conference Working Group on Electronic Document Retention and Production. He also serves as a member of Squire Sanders’ E-Discovery & Data Management Team. He is an ACEDS

Certified E-Discovery Specialist and since 2011 has served on the Executive Board of the Association of Litigation Support Professionals (ALSP). He is also a Planning Committee Chair for the Georgetown Law e-Discovery Practice Support program.

**Sean P. Foley, Esq.** is Project Manager, ProSearch Strategies, Inc. His experience is in consulting on matters of information governance, case preparation and document management, and in leading large scale automated document review engagements on behalf of Fortune 500 companies. A licensed California attorney, Mr. Foley spent five years handling all aspects of complex case litigation up to and including trial and binding arbitration; motion practice on discovery issues; arguing dispositive motions; and participation in early resolution and mediation of cases. Prior to joining ProSearch, he was Manager of Research and Implementation at H5, where he managed domestic and offshore review teams.

**Please email the panelists at [ccrowley@crowleylawoffice.com](mailto:ccrowley@crowleylawoffice.com), [stephen.goldstein@squiresanders.com](mailto:stephen.goldstein@squiresanders.com) or [sfoley@prosearch.us](mailto:sfoley@prosearch.us) with questions about this panel discussion.**

**any difference between this and what people refer to as technology-assisted review or computer-assisted review?**

**Crowley:** Predictive coding involves the use of software that has the ability to take determinations made by a human with respect to document relevance and then to leverage those determinations across a larger body of documents. Predictive coding focuses the initial review or training effort on a subset of documents and the determinations of document relevance on these documents are then extrapolated across the entire population. Predictive coding is one aspect of a larger umbrella term – technology-assisted review or computer-assisted review.

**Editor: In the view of the panelists, what is the significance of the Opinion?**

**Goldstein:** There have been a number of corporations and law firms, and ours is included, that have been using predictive coding for some time. That said, a lot of people may have been observing from the sidelines, waiting for a federal judicial opinion to provide greater certainty about the defensibility of predictive coding. It's unclear how many of them will now jump in, but the decision certainly pushes the topic to mainstream conversation among corporate clients and outside counsel.

**Foley:** As Steve mentioned, many forward-thinking corporations and law firms confronted with huge volumes of data have already adopted predictive-coding technology, particularly in their large-scale reviews. The significance of the Opinion was perhaps more of a watershed moment as it allows e-discovery vendors to expand the market for their products.

The Opinion is not so much the starting gun but rather legitimizes a process that has been utilized for some time. It's worth noting that the Opinion was qualified in the sense that it states that the particular process that was used in the case was acceptable to the judge under the circumstances of that case.

**Crowley:** It's a watershed moment in one respect. Although we've had implicit endorsement prior to this from Judge Facciola and Judge Grimm, Judge Peck provided the first written judicial opinion that specifically endorses the use of predictive coding.

**Editor: Is it likely that someone will express a contrary opinion?**

**Crowley:** No. Independent studies have demonstrated the efficacy of predictive coding, as was recognized in the Opinion.

**Editor: Have issues of defensibility become a blip in the history of predictive coding?**

**Goldstein:** Defensibility is still a concern. Judge Peck is quite explicit in saying you have to be able to prove a valid process. You have to be able to demonstrate your control, your testing, your validation, and your general understanding of the method so that you can show that it has produced an appropriate level of recall and precision.

**Foley:** As far as defensibility is concerned, it is important that the circumstances of the case be right. The Opinion was based on a situation involving a generally cooperative attitude on the part of the litigants as reflected in a fairly open process of sharing seed documents and methodologies. I'm not sure that's a standard that all litigants will want to follow. There is a case to be made to keeping all of your seed documents private. Additional issues of defensibility are raised where there's a black box or other complex technology. While Judge Peck has said that the process dealt with in the Opinion was defensible, there are different processes and enough variety in the way they are deployed to assure that defensibility will continue to be an issue.

**Goldstein:** The reading of the transcript of the hearing that led to the Opinion is instructive for what could happen if you're not prepared to defend the process.

**Editor: What are some of the defensibility issues that could be raised in a case of this kind?**

**Crowley:** The areas of attack that could present themselves are now focused on the process employed as opposed to the technology. So the defensibility issues would relate to the transparency of process, the reasonableness of the process, the quality assurance that is built into the process, and the ability to report on the metrics involved. It is now essen-

tial to have an appropriately designed, defensible process.

**Editor: To what extent is the quality of the senior lawyer who might be training the black box an area of possible attack?**

**Goldstein:** It's probably one of the larger ones. If these tools do what they are supposed to be doing, they will mimic the training they receive, and so if they're trained by a second-year associate, it will probably reflect that, as opposed to being trained by a senior-level associate or partner.

**Crowley:** The software is only as good as the information with which it is provided. Unless the lawyer providing that information has an intimate understanding of what constitutes relevance in a particular matter, the software will not be able to make valid determinations as to relevance.

**Editor: Doesn't this put a lot of pressure on the side that's not as adept at using this process to bring itself up to speed?**

**Foley:** We now have the Opinion to rely on and the technology is fairly well documented – it's been studied a great deal. If one side says using this technology is a trusted way to get results that are at least as good and likely much better than doing a manual review, it shifts the burden to the other side to say no, we don't accept this technology and here's why you have to go ahead and spend insane amounts of money to do manual review. The side that's questioning the technology is compelled to have the expertise to do so in a competent manner. The fact that it may be perceived as confusing or complex is not a sufficient justification to impose the massive cost of manual review on any litigant.

**Crowley:** If you look at traditional, linear manual review, you have a senior attorney who's familiar with the issues in the case, who attempts to communicate that knowledge to a whole swath of junior attorneys. That information transfer is subject to misinterpretation and misunderstanding. The advantage we have with predictive coding is that a senior attorney unequivocally communicates his or her understanding of relevance to the soft-

ware with no misunderstanding, no opportunity for misinterpretation, and those high-level determinations are then extrapolated across the population to ensure consistent relevancy determinations across a large body of documents. It's a major advantage.

**Editor: Does the Opinion stand for the proposition that objections to predictive coding are dead and buried under certain circumstances?**

**Crowley:** Judge Peck made clear that he considers the idea that you can attack software that does what it purports to do to be an issue that's dead and buried. His opinion also made it clear that issues relating to the defensibility of the particular processes in which it is used will continue to arise.

**Editor: Is predictive coding yet another industry fad that will be forgotten in a year or two?**

**Foley:** Absolutely not. The proliferation of data and the ability to inexpensively store really large amounts of data means that some principled manner of wading through these volumes is going to be required. Predictive coding is the way to do that.

**Editor: What additional benefits does predictive coding provide?**

**Crowley:** It lowers the cost to the clients and ensures that the expertise they are paying for at the senior level is being properly deployed. We've seen a fundamental shift in the way clients approach law firm services because they're no longer willing to pay for junior-level attorneys to do document review. Rather than paying for young associates that do not have a tremendous amount of expertise or experience to review documents, what clients are now getting, via predictive coding, are senior-level attorneys promptly providing their insights.

**Goldstein:** The technology provides a number of valuable benefits. Not only does it reduce the cost of discovery, it also allows us to budget the cost of the discovery much more accurately. When you look at the expense of reviewing and producing a large document collection, we're able to very closely estimate the billable review time that might be

involved using predictive coding tools and workflows.

Further, because relevancy values can be assigned to each document in the corpus within a short time period, we can quickly bubble up to the top the most relevant documents in a large population.

This allows the attorneys to rapidly gain an understanding of what the case looks like – and I mean within days of starting the process. In a linear type of human review process, you have a series of junior-level people looking through mountains of documents that slowly trickle up to the decision makers – the partners and the client. It may be weeks and even months before they see the hot documents. This early insight can help the team develop a well-informed strategy, or a fight /flee decision. This is particularly useful in government investigations.

Mainly though, in a predictive coding workflow, those lawyers that are involved in the software-training process are educated about a case within the first couple days of the training process in a way that would normally have taken months. They're seeing the critical content hands on. This accelerates their acquisition of knowledge of the material and their understanding of the issues of the case. It's been a fascinating shift in how the senior attorneys learn about a case.

**Editor: Can you share your own experiences in using predictive coding technology?**

**Goldstein:** We've been using this process for two years now, and it's a standard part of our workflow for large matters. Any case involving more than 40,000 or 50,000 documents is a potential candidate, and it's been an extremely effective experience. We typically achieve a 65-85 percent reduction in the data volume. A recent example is a matter in which we started with 800,000 documents, on which we trained the predictive coding system. After a thorough validation, and in consultation with the client, we were able to exclude more than 650,000 documents from further attorney review.

**Foley:** We have employed predictive coding technology on matters with data populations exceeding 10 million documents and have been very pleased with the results.

**Editor: Does Judge Peck provide guidelines about how a predictive coding project should be conducted?**

**Goldstein:** He is looking for some cooperation, which may be hard for some litigants to adjust to. In particular, he suggests that the other side should be privy to the documents that were used to train the machine. He doesn't say that you have to explain to the other side why you decided to tag those documents as relevant or not, but he does suggest there be some transparency.

**Crowley:** All I would add is that what he really is emphasizing in his opinion is the importance of the process employed far more than the specific technology utilized.

**Foley:** Judge Peck puts emphasis on process transparency as opposed to content transparency. We believe that sharing of the seed documents, even the non-privileged documents that were used to train the system, should not be mandated. One of the takeaways is that having an agreed-upon protocol and having transparency in the process is key.

**Editor: Over the last couple of years, even since the emergence of the technology, discussion about defensibility has been focused largely around the underlying black box algorithms. How does Judge Peck's opinion weigh in on that debate?**

**Crowley:** I would just like to address the concept that the technology is a black box. We've been dealing with black boxes for decades. In traditional linear review, the black boxes are the minds of the junior attorneys who reviewed the documents. There is no transparency with respect to how these attorneys determined relevance. What Judge Peck does is recognize the efficacy of predictive coding for making that determination. He has moved the discussion beyond the efficacy of the algorithm to the defensibility of the process employed in utilizing the software to make the relevance determination.

**Goldstein:** What's interesting is that the parties seemed to focus on the black box issue to the point that they were throwing rocks at each other about how the technology works. It was the judge who con-



tinually steered them back to the process.

**Crowley:** We have moved beyond the technology issue now to discussions of the extent to which the parties should be transparent and how one should focus on issues of defensibility in the process employed rather than the specific algorithms employed by a particular piece of technology.

**Foley:** I agree. You can say we are using this particular software, but the potential infirmities in the review remain, just as improper supervision of contract attorneys is going to lead to inaccuracies in the review. If you do not train and supervise the system properly, you likewise are going to have issues. You can deploy any piece of software, but are you testing it and doing enough quality control? Predictive coding technology only gets you so far. There are a lot of other processes and techniques that need to accompany it.

**Editor:** Is predictive coding the exclusive province of technology aficionados, or is this a technology that can be adopted for mass use?

**Goldstein:** It's becoming more accessible. But my conclusion from the Opinion is that you have to invest some amount of time in understanding these tools and the process. It's not just a service you can pay for. There has to be some ownership and some understanding by counsel.

**Foley:** The vendor needs to explain the process to counsel before counsel can go to the court and affirmatively state that they have done a competent review. Counsel needs to be able to understand and verbalize and discuss what it is that they've done, and it's really important for anyone promoting predictive coding technology to be able to communicate effectively the ways you are going about

quality control and show how you are leveraging the software. Buying a license for the software doesn't mean anything. How are litigants using it? How are they supervising the system, and how are they checking the quality of the review?

If the lawyer doesn't understand the underlying facts of the case, he or she is not going to train the system well. You have to get subject matter experts involved whether or not they are fans of the technology. To train the system effectively you need the most knowledgeable people involved, because you're impacting relevance determinations for tens of thousands of documents every time you click through a training set.

Using predictive coding technology can't be the exclusive province of technology aficionados. It needs to be adopted and used even by those who are more resistant. The people who understand the facts and the procedural posture of the case need to be involved in training the system. Getting experts involved in the process is critical to ensuring the quality of the results.

**Editor:** Where do you think predictive coding will take us five years from now?

**Crowley:** I think it will allow us to return to having highly experienced lawyers focus on the issues in the case more swiftly and effectively. In five years' time, we'll take it for granted that you can gain a quick understanding and assessment of the information that you have available for a particular case. Predictive coding will allow us to deploy the expertise of senior counsel to benefit clients rather than relying on potentially divergent relevance determinations made by hundreds of human reviewers.

**Foley:** In five years' time, predictive coding will be universally used in cases

involving significant amounts of ESI. People will look back and say, I can't believe we weren't using predictive coding earlier. That is the reaction we get from our clients. Once they're involved in the process, and once they actually see the results, they become believers, and then some of them become evangelists of the process and of using predictive coding. In five years people will look back and say wow, we were doing it wrong.

**Goldstein:** If you look at some trends or disruptive events over the last 15 years, there was a time when large firms wouldn't use contract attorneys and now that's become widely accepted. Then we had to get through the use of keywords, which was hard to accept and adopt, but eventually became commonplace. I think predictive coding will become part of a routine process within the next three to five years. Some firms will get there fast, while those who are more risk averse will take longer. The technology may even force a change in the business model of firms who rely on large volumes of document review. Clients and insurers will eventually demand the reliability and cost savings of these tools.

**Foley:** It's going to allow lawyers to go back to lawyering instead of worrying about getting through huge volumes of data. They'll be able to ascribe meaning and insights to the document population instead of seeing it as a huge volume of documents that needs to be plowed through. With predictive coding, you have the ability to gain insights into your case very early and to achieve a better understanding of the facts of the case and the sort of data that has been collected. Predictive coding to a certain degree allows lawyers to use their legal expertise in preparing their cases instead of managing data; in other words, lawyers will go back to being lawyers.