



LJN's

LEGAL TECH Newsletter®

An ALM Publication

Volume 31, Number 6 • October 2014

e-Discovery Law

Court Praises Predictive Coding, Then Rejects It

By H. Christopher Boehning and Daniel J. Toal

Over the past few years, much ink has been spilled by judges, commentators, and e-discovery service providers opining on the merits and drawbacks of predictive coding. (As we noted in our article, "Seed Set' Documents Should Not Be Discoverable," 251 *NYLJ* 2 (Feb. 4, 2014), in which we provide background on predictive coding and review a number of judicial decisions on the topic, predictive coding, also often referred to as technology assisted review, is the use of computer-generated algorithms to supplement and extend the work of human reviewers in the discovery document review process.) Topics have included when predictive coding should and should not be used, which of the many competing predictive coding technologies and processes are best, whether predictive coding is more accurate than human review, and exactly how, if at all, predictive coding should be incorporated into discovery protocols.

In one of the best-known court battles over predictive coding, *Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012), the
continued on page 6

ESI Technologies

It's Not One or the Other, But Both

By Michael Conner and Jeffrey Teso

In the world of e-discovery, we frequently become caught up in "latest and greatest" cutting edge technologies, often forgoing the previous. It wasn't that long ago when early case assessment (ECA) tools were all the rage. Practitioners were discussing how removing large swaths of data from a document data universe could provide enormous cost savings, as well as provide greater insight leading to more transparent project scoping and budgeting. As many of us were settling on technology platforms and forming case strategies and protocols utilizing ECA software, predictive coding (PC) and technology assisted review (TAR) software entered the market. The latter almost too quickly seemed to replace the former. The fact is, both technologies used on their own can provide great benefit, however, they should not be considered mutually exclusive. The advantages of creating an ESI strategy that includes the use of each, ECA and PC or TAR has numerous benefits that should not be overlooked.

IN WITH THE NEW

Recent years have brought significant advances in legal search and review technologies, offering practitioners a new and enhanced set of tools to assist in their efforts to combat the continuous growth in data volumes and the effort and cost associated with responding to discovery, all while satisfying their clients' obligations in a defensible manner. The most important of these advances includes a category of technologies and processes, commonly known as technology assisted review (TAR), computer assisted review, automated review or predictive coding. For simplicity, this group of technologies is collectively referred to as "TAR-based technologies" throughout this article. Due to their ability to organize and dispose of large volumes of documents with limited human interaction, many practitioners view TAR-based technologies as the tool of choice for significantly reducing

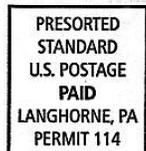
continued on page 2

In This Issue

ESI Technologies.. 1

Court Praises Predictive Coding, Then Rejects It 1

Efficient Review In a Time-Sensitive Government Investigation..... 3



ESI

continued from page 1

the overall effort and cost associated with analyzing and organizing large sets of documents.

Most TAR-based technologies incorporate a form of machine learning that leverages the review decisions (typically, relevant or non-relevant) made by subject matter experts (SME) on a small number of reviewed example documents, to automatically extrapolate those decisions to the remaining set of unreviewed documents. Although this approach can vary based on the specific software used, the basic steps followed by practitioners following a TAR-based workflow typically involves the following three activities: 1) training of the software's machine learning algorithm through the manual review of a small number of example documents; 2) application of the software's learning (machine categorization), perfected through iterative rounds of training and categorization; and 3) validation of the software's automatic categorization through the manual review of a statistically relevant random sample of machine categorized documents.

OUT WITH THE OLD:

EARLY CASE ASSESSMENT (ECA)

Early case assessment (ECA) applications are another group of litigation support technologies, long predating TAR-based solutions, which offer practitioners another effective tool in satisfying their primary discovery objectives. These technologies include a group of soft-

Michael Conner and **Jeffrey Teso** are Managing Directors at the global professional services firm Alvarez & Marsal. Based in San Francisco and Dallas, respectively, they serve in the firm's Global Forensic and Dispute Services practice. Conner is an e-discovery expert who focuses on the collection, processing, and production of electronic data involved in litigation. Teso is a former practicing attorney who now heads the firm's managed document review offering.

ware products containing advanced text and metadata search and visual analytics features that allow users to evaluate the objective and relational characteristics of large data sets. ECA tools allow users to organize and filter documents by both content and meta data attributes, including name, e-mail address, domain, date, file type, communicant grouping, text cloud, etc., giving counsel the ability to quickly and defensibly include presumptively relevant content or exclude obviously irrelevant content with minimal effort or risk, and thereby reduce the downstream cost associated with wasteful effort. While once available only in stand-alone applications, ECA features and functionality have now been incorporated into many traditional litigation support technologies, including litigation hold, ESI processing and document review software.

Conventional wisdom generally recognizes ECA tools as being ideally suited for initial case activities, including specifically the investigation and culling of starting document sets, an act typically performed after processing and prior to review to defensibly reduce downstream effort and cost. ECA tools are also commonly used to gain an early understanding of the facts of a matter, as well as to make informed decisions regarding scope, risk and overall response strategy.

ECA OR TAR? YES

When performing the search and analysis of ESI required to respond to production requests and internal investigations, counsel's activities are typically governed by three primary objectives: 1) gaining an understanding of the facts central to the litigation or investigation; 2) ensuring that his or her client satisfies its production obligations in a reasonable and defensible manner; and 3) accomplishing both of the above with the least amount of effort and cost. ECA and TAR-based technologies each have a role to play in meeting these objectives, and both can — and should — be used to produce optimal results.

continued on page 5

LJN's

LEGAL TECH[®] Newsletter

EDITOR-IN-CHIEF Adam Schlagman, Esq.
EDITORIAL DIRECTOR Wendy Kaplan Stavino
MANAGING EDITOR Steven Salkin, Esq.
MARKETING MANAGER Jeannine Kennedy
GRAPHIC DESIGNER Evelyn Fernandez

BOARD OF EDITORS

RICHARD C. BELTHOFF JR. Wachovia Corp.
Charlotte, NC

BRETT BURNEY Burney Consultants
Cleveland, OH

JEFFERY M. DUNCAN Brinks Hofer Gilson & Lione
Chicago

TOM GELBMANN Gelbmann & Associates
St. Paul, MN

JOHN GREEN Baker, Donelson, Berman,
Caldwell & Berkowitz, PC
Memphis, TN

JUSTIN HECTUS Keesal, Young & Logan
Long Beach, CA

RICHARD K. HERRMANN Morris James LLP
Wilmington, DE

SUE HUGHES PayneGroup
Seattle

ERIC HUNTER Bradford & Barthel
Sacramento, CA

ARI KAPLAN Ari Kaplan Advisors
New York

MARC LAURITSEN Capstone Practice Systems
Harvard, MA

RICHARD D. LUTKUS Seyfarth Shaw LLP
Chicago

JIM G. MITCHELL Huron Consulting Group LLC
Chicago

DAVID C. REYMANN Parr Waddoups Brown
Gee & Loveless
Salt Lake City, UT

SETH A. RIERSON FTI Consulting, Inc.
Chicago

G. CHRISTOPHER RITTER The Focal Point LLC
Oakland, CA

JOEL B. ROTHMAN Arnstein & Lehr LLP
West Palm Beach, FL

DONNA SEYLE Law Practice Strategy
Santa Cruz, CA

GEORGE J. SOCHA JR. SochaConsulting LLC
St. Paul, MN

CHRIS SPIZZIRRI Travelers
Hartford, CT

JOHN J. SROKA Duane Morris LLP
Philadelphia

DANNY THANKACHAN Thompson & Knight LLP
Dallas

BEN WEINBERGER Bond Pearce LLP
Bristol, UK

LJN's Legal Tech Newsletter® (ISSN 0738-0186) is published by Law Journal Newsletters, a division of ALM. © 2014 ALM Media, LLC. All rights reserved. No reproduction of any portion of this issue is allowed without written permission from the publisher.

Telephone: (877) 256-2472;
Editorial e-mail: ssalkin@alm.com
Circulation e-mail: customer@alm.com
Reprints: www.almreprints.com

POSTMASTER: Send address changes to:
ALM
120 Broadway, New York, NY 10021

Published Monthly by:
Law Journal Newsletters
1617 JFK Boulevard, Suite 1750, Philadelphia, PA 19103
www.ljnonline.com



Efficient Review In a Time-Sensitive Government Investigation

Developing and Implementing Cost-Effective Processes

By Sanjay Manocha

Over the past 10 years, government investigations have become increasingly sophisticated in analyzing electronically stored information (ESI). Federal executive departments and agencies have made substantial investments in advanced analytical systems that help investigators and prosecutors filter voluminous amounts of incoming ESI to quickly focus on items of particular interest and relevance to an investigation. These systems, once almost magical in the speed and depth of their analysis, are now commonplace.

Companies and organizations responding to Civil Investigative Demands (CIDs) and other government requests for information must recognize that the information provided will be analyzed using these powerful tools. Significant documents in a voluminous production that previously might have been overlooked will now likely be discovered. Even more importantly, prosecutorial data analysis may reveal documents to the investigators that, in the absence of similar capabilities, a target company may not be aware it was producing.

Recently, RVM Enterprises, Inc. (RVM) worked with an AmLaw 100 firm and one of the world's largest corporations to respond to voluminous and time-sensitive requests made in connection with an inves-

Sanjay Manocha oversees implementation of advanced analytics and predictive coding technologies in discovery practice at RVM Enterprises, Inc. Prior to RVM, Manocha was CEO of N-Tier Discovery, a discovery analytics consulting firm, and practiced law, specializing in complex commercial litigation and regulatory investigations.

tigation conducted by the United States Department of Justice (DOJ). The DOJ sought both the production of documents and fact witnesses for depositions, and the law firm and its client had less than three weeks to analyze over five million potentially responsive e-mail messages and other documents already produced to the DOJ.

The client initially intended to use a traditional keyword search approach to identify "hot" documents for use in witness interviews, but a traditional document review method could never have met this tight deadline. Equally important, the client's legal response team understood that the DOJ investigative team was using some of the DOJ's more robust ESI analytical tools to "data mine" the target company's document production, and the team wanted to do everything it could to ensure that the DOJ's analysis did not turn up any documents that had not already been evaluated by the client's response team.

APPROACH

As requested by the client, we started the project by indexing the collection and running initial keyword searches developed by the client. It is well documented that keyword search queries are both over-inclusive and under-inclusive; they generate false positives that aren't genuinely relevant to the matter being investigated at the same time that they do not find many of the relevant documents that do not contain *any* of the keyword search terms used. Some studies have demonstrated that traditional keyword searches alone may miss up to 80% of the relevant documents in a collection. See, *Da Silva Moore v. Publicis Groupe*, 2012 U.S. Dist. LEXIS 23350 at 19 (S.D.N.Y. Feb. 24, 2012) (citing, *inter alia*, David L. Blair & M. E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," 28 *Comm. ACM* 289 (1985)).

Recognizing that keyword search alone was insufficient for this project and its compressed timeframe, we suggested that a multi-modal, multi-platform approach would

best meet the needs of the client's response team. Traditional methodologies, such as standard keyword search, e-mail threading, near-duplicate identification and concept clustering would, in each case, help winnow down the overwhelming volume of potentially responsive documents to a more manageable level, but no single tool could do the whole job on its own. In addition, the team identified the need to deploy advanced predictive coding technology.

We also explained to the client that the order in which tools were deployed would play an important role in determining how efficient culling the mountain of documents would be, while simultaneously flagging potentially significant documents for closer analysis. Tools can be deployed in different order within a workflow depending on client objectives; here, RVM advised and the client agreed to an approach that first focused on triaging relevant data before eliminating redundant documents for concept clustering. As a final step, RVM leveraged conceptual search to enhance the ultimate attorney review and investigation of the responsive documents.

To begin triaging the document collection, we took all documents, along with known documents of interest that had already been identified by the legal team, and deployed Equivio's Relevance predictive coding technology to begin to filter and prioritize further analytical work.

Legal team members used Equivio Relevance to directly view sample document sets from the collection and rate the relative importance of those documents. Each sample set categorized by the legal response team was then used to "train" the Equivio Relevance tool. Relevance would then update its document rankings by relative importance and spawn an additional sample set for attorney review. Results of this and subsequent rounds of Equivio Relevance training would continue to improve the system's accuracy until

continued on page 4

Equivio

continued from page 3

the Relevance rankings closely matched the legal team's understanding of the substance and relevance of the documents.

Here, Relevance training proceeded quickly, and the legal team needed only three iterations of Relevance system training before Equivio Relevance was able to reliably rank documents by relative importance.

Once the documents were ranked in terms of relative importance, it was important to find the appropriate dividing point between likely relevant and likely irrelevant documents. The RVM team used sampling and other techniques to help the legal team choose an appropriate cut-off point for this review, based on both objective and subjective analytical factors. Based on the importance of the investigation, the legal response team erred on the side of caution, choosing a somewhat lower cut-off point to ensure that as many responsive documents as possible would be passed to the next stage of analysis.

To further filter and prioritize the documents above the Relevance cut-off point, RVM next deployed Equivio's near-duplication and e-mail threading tools to identify documents that were substantially similar to each other or, in the case of e-mail threading, all messages generated from a common strand of e-mail "conversations."

This analytical approach to near-duplicate identification gave RVM and the client's legal response team the ability to group document drafts comprising largely duplicate content, and in so doing reduce the number of documents requiring individual attorney review.

Similarly, Equivio's e-mail thread analysis identified "inclusive" e-mail messages that included the text of all prior messages in a thread. Therefore, reviewing a single inclusive e-mail message would eliminate the need for an attorney to review all other messages in a thread. Where e-mail conversations fragmented into multiple conversations,

the system identifies multiple inclusive e-mail messages to account for message forks. From the outset, our team knew that the client's legal response team would ultimately conduct its final relevance review and document production using kCura's Relativity platform. We loaded the documents into a Relativity workspace and then leveraged Relativity's own functionality to further prioritize the review via clustering analytics applied to Equivio-identified pivots and inclusive documents.

Combined with Equivio Relevance rankings, this permitted the legal team to prioritize clusters containing known high-value documents and these clusters were immediately reviewed by the legal team for substance.

Equivio Relevance rankings were also key to identifying clusters of documents that contained both high-ranking and low-ranking documents. Why did these clusters exist? Were there additional concepts or vocabulary that that hadn't previously been recognized? Again, cross-referencing the analysis from two disparate systems permitted immediate, reproducible, and highly defensible quality control that was constantly being used to refine the eyes-on review and re-prioritize document clusters.

Systematic, substantive review also required consistent rules and results. To meet this requirement for consistency, the legal response team crafted a detailed written review protocol that defined the review categorization methodology based on initial analysis derived from briefings by the client, known high-value documents that were the results of keyword search, and Equivio Relevance training.

The legal team's protocol also provided flexible "rules of engagement" for the collection. Rather than focusing only on defined clusters or review batches, reviewers were encouraged to use the platform's "find similar documents" and "expand terms" functionality whenever an unexpected document of interest was identified. When a review team member found a particularly significant document, Equivio near-dupli-

cate or e-mail threading would help the team member see the entire context for the document, whether or not all of these documents were in the original cluster where the first document was found.

Behind the scenes of the legal team review, the RVM team regularly collected the results of the legal team's eyes-on document review and used their review decisions to update the concept clusters — adding new clusters and re-categorizing the priority of existing clusters based on the ongoing analysis of the legal team.

Having this behind-the-scenes review in place was significant because it served as an iterative process to continually improve the organization of the documents with new information as it was discovered. We partnered with the legal review team by speaking with them on a daily basis to keep all members of the response team on the same page.

RESULTS

How long did this entire process take?

At the outset, the legal team of five associates faced the Herculean task of analyzing five million documents in three weeks. Departing from tradition, and using RVM's Structured Review (RSR) approach, they completed their task of first pass review, from start to finish, in just three weeks (17 days).

During that 17-day period, the legal response team, using RVM's multi-platform approach, reduced the initial population of five million documents to 400,000 documents potentially requiring eyes-on review by the legal team — a 92% reduction.

Once the legal team began its substantive document review, RVM's heavy use of analytics permitted the remaining documents to be further prioritized for review, giving the team great confidence in the review results without requiring individual review of the full 400,000 documents.

Analysis at the end of the project demonstrated the noteworthy discovery that, if not for the

continued on page 8

ESI

continued from page 2

The traditional view of many attorneys and litigation support professionals is to draw a clear distinction between the value and appropriate use of ECA tools and TAR-based technologies, with each relegated to a specific and rigid role along the continuum of activities undertaken to respond to production requests or internal investigations. In practice, workflows that combine ECA tools and TAR-based technologies can often yield vastly improved results versus the use of a single technology. Indeed, in most cases optimal results can only be achieved when both tools are intelligently and flexibly used during each stage of the process.

USING AN OLD TOOL TO IMPROVE A NEW ONE

An example of the smart integration of ECA and TAR-based technologies — often overlooked by practitioners due to fundamental misconceptions about each tool's exclusivity and suitability — can be found in execution of a traditional TAR-based review workflow. When used smartly, ECA tools can improve counsel's ability to enjoy the full benefits of advanced TAR-based workflows.

Most TAR software includes features and functionality that allows users to perform all three of the traditional steps of a TAR workflow within the TAR application itself, without the need to incorporate additional technologies. Further, many attorneys and litigation support professionals believe that, in order to work properly, these advanced tools must be used exclusive of other technologies. As a result, when executing TAR-based workflows, most practitioners restrict their activities to a single application and the specific process steps suggested by the software developer offering the technology.

ECA to Improve Data Set Richness

The subject of whether TAR-based workflows require the analysis of fully inclusive, uncultured document

sets is much debated, and to date no generally accepted best practice has emerged. When following TAR-based review workflows, many legal professionals elect to forego initial culling efforts altogether, believing optimal results can only be achieved when the software's training involves the largest and most inclusive starting set of non-culled documents. In many cases, this initial, post processing universe of documents can be extremely large, due to the overly broad preservation and collection activities that are frequently undertaken by counsel to ensure defensibility. When the starting universe has low richness, meaning the percentage of relevant content within the universe is relatively small, counsel's efforts to categorize the full universe using a TAR-based workflow can often require several rounds of training and categorization, at significant effort and cost.

One approach to improve the retrieval quality of TAR software is to incorporate ECA-based technologies to increase the richness of the universe to be categorized by removing obviously irrelevant content. Further, in practice, counsel's ability to obtain value from TAR-based workflows appears to be unaffected by the initial culling of irrelevant content prior to the software's training and categorization. Because most document sets contain a disproportionately higher percentage of irrelevant content to relevant content, even after standard keyword and date filtering, the use of ECA tools to remove additional, obviously irrelevant content frequently does little to impact the TAR software's ability to categorize the remaining documents. As such, combining upfront ECA-based culling with subsequent TAR-based categorization represents an effective approach for driving greater cost efficiency into the use of TAR-based workflows.

ECA to Improve TAR Training

ECA tools can also play an important role in the training of TAR-based technologies. While there remains significant debate regarding the best method for selecting the

example documents to be reviewed to train the software, retrieval quality is generally improved when the SMEs training the system are able to consider a diverse and representative cross section of both relevant and irrelevant content. Identifying that cross section is typically accomplished through random or judgmental sampling of the starting document universe using basic keyword or other text-based searches within the TAR software platform. For practitioners choosing to seed their training sets through judgmental sampling, performing purposeful, focused searches using an ECA technology can serve as an effective tool for uncovering a richer cross section of representative content. The advanced search and visual analytics features of these tools allow users to perform more focused queries to uncover a broader and visually verifiable variety of both relevant and non-relevant content segments, which can in turn lead to faster and improved TAR categorization.

ECA to Improve TAR Validation

One of the more challenging steps of standard TAR-based workflows involves the final stage, when counsel is required to validate the quality of the software's machine-based categorization. This validation is typically achieved through the manual and iterative review by SMEs of a statistically relevant random sample of documents categorized automatically by the software. If that review demonstrates a repeated and acceptably low occurrence of error in the software's machine-based decisions, counsel will frequently suspend further validation efforts and consider the process complete. If the error rate following validation review remains unacceptably high, counsel will typically re-categorize using the additional coding decisions revealed during this review and will repeat the process until the

continued on page 6



ESI

continued from page 5

error rate falls within a repeatable and acceptable range.

Often, counsel's primary objective in validating the reliability of the TAR software's categorization is to demonstrate the defensibility of foregoing the effort and cost associated with manual review of validated non-relevant content, which requires measurement of error rate. In practice, achieving an error rate low enough to defensibly rely on the TAR software's categorization can be difficult, and often requires several rounds of validation review and re-categorization. In some cases, arriving at an acceptable error rate might ultimately not be achievable,

regardless of the number of validation and re-categorization rounds performed. This can often result with large collections of documents machine categorized as non-relevant due to general prevalence of irrelevant content in those documents, but which also contain at least some portion of relevant content.

While statistical sampling is generally considered an acceptable approach for measuring the results of a TAR software's machine categorization, the method relies on the sampling of a relatively small amount of content randomly selected from a large universe of documents. This process can often lead to frustration when multiple rounds of validation review fail to generate acceptable error rate metrics. When faced with this

outcome, many attorneys will simply abandon the sampling approach altogether and elect to perform a full validation through traditional, costly, linear review. The use of ECA tools during the validation stage of TAR-based workflows can often help counsel avoid this result. Leveraging the more focused and visually verifiable search methods available in ECA technologies, counsel can surgically target and extract known relevant content from larger universe of machine categorized, non-relevant documents. By identifying the prevalent characteristics of the known relevant universe through an ECA platform, counsel can use those characteristics, including most prevalent concepts and keywords, relevant

continued on page 8

Predictive Coding

continued from page 1

court concluded that a judge could incorporate predictive coding into electronically stored information (ESI) protocols over the objection of one of the parties. But what happens when the parties have already agreed to an ESI protocol, which does not include predictive coding, and one party wants to change this protocol, over the objection of the other party, by incorporating predictive coding? Can one party unilaterally alter the method of discovery, or does the agreed-upon ESI protocol take precedence?

H. Christopher Boehning and **Daniel J. Toal** are partners in the Litigation Department of Paul, Weiss, Rifkind, Wharton & Garrison LLP. Boehning's practice includes complex commercial and civil litigation matters, insurance counseling and litigation, regulatory inquiries, internal investigations and international arbitrations. He can be reached at cboehning@paulweiss.com. Toal's practice concentrates in securities litigation, employment law, environmental litigation, and complex commercial disputes. He can be reached at dtoal@paulweiss.com.

This is the question addressed by the court in *Progressive Casualty Ins. v. Delaney*, No. 2:11-cv-00678, 2014 WL 3563467 (D. Nev. July 18, 2014). In answering this question, the court sheds light on the reasons why parties have been reluctant to accept predictive coding — the need for cooperation and transparency with one's adversary and the resulting risks of this cooperation — and highlights a key debate over best practices — whether search terms can be used to first limit the universe of documents before predictive coding is employed.

Progressive Casualty Ins. v. Delaney

In this litigation, Progressive was seeking a declaratory judgment that its Directors and Officers Company Liability Insurance Policy for Financial Institutions issued to various failed banks did not provide coverage for lawsuits initiated by the Federal Deposit Insurance Corporation (FDIC) against former directors, officers, and employees of the failed banks.

In the course of this litigation, on Oct. 24, 2013, Progressive and the FDIC agreed to a joint ESI protocol (ESI protocol) which, *inter alia*, prescribed the process for searching, reviewing and producing ESI. This ESI protocol, which followed the

same template used by the parties in five other related lawsuits, was formalized as a court order. *Id.* at 5.

Progressive collected 1.8 million documents in its discovery efforts. The parties agreed to a set of search terms to filter this full set of documents, which the FDIC "spent months narrowing at Progressive's insistence" and which yielded 565,000 potentially relevant documents. *Id.* Pursuant to the ESI protocol, Progressive could either turn over all 565,000 documents without reviewing them for responsiveness, or it could choose to undertake a review and withhold any privileged or non-responsive documents. And, as the court noted, the "ESI protocol addressed inadvertent disclosure of privileged materials and contained clawback provisions pursuant to Fed. R. Civ. P. 26(b)(5)(B) and Fed. R. Evid. 502(d)." *Id.* at 7.

Progressive was supposed to produce documents on a rolling basis. An initial set of documents was due to be produced by the end of September 2013. When Progressive

continued on page 7

The publisher of this newsletter is not engaged in rendering legal, accounting, financial, investment advisory or other professional services, and this publication is not meant to constitute legal, accounting, financial, investment advisory or other professional advice. If legal, financial, investment advisory or other professional assistance is required, the services of a competent professional person should be sought.

Predictive Coding

continued from page 6

missed that deadline, it told the FDIC that it would produce an initial set of documents in mid-October. Again, Progressive did not produce any documents. As of Dec. 27, 2013, the FDIC had yet to receive the vast majority of Progressive's production and had not received any production of ESI, even though discovery was set to close less than three months later.

This delay was due to Progressive's choice, as was specified as an allowable option in the ESI protocol, to review the 565,000 documents using a traditional process of having dedicated review attorneys manually review each document for responsiveness. After eight contract attorneys had reviewed 125,000 of the documents, Progressive, recognizing that discovery deadlines were looming, determined that the manual review was too time intensive and expensive and began exploring alternative options. Progressive turned to predictive coding to speed up its review. Notably, however, predictive coding was never part of the initially agreed-upon ESI protocol and Progressive never informed the court or the FDIC that it was employing predictive coding; only after the FDIC filed its motion to compel on Dec. 27, 2013 did the parties meet and confer regarding the possibility of its use. *Id.* at 4.

During meet-and-confers in January 2014, the parties could not reach a resolution regarding the use of predictive coding. This was in part due to Progressive's reluctance to provide detailed information about how it planned to apply predictive coding. Progressive's counsel refused to provide such details based on a belief that the FDIC's request was part of a "pattern of making unreasonable discovery demands." *Id.* at 9. After the meet-and-confers, the FDIC opposed Progressive's proposed use of predictive coding, arguing that it would complicate discovery and lead to numerous satellite disputes. To the extent the

court was inclined to allow predictive coding, however, the FDIC set forth an alternative proposal. While Progressive wanted to apply predictive coding only to the 565,000 documents that resulted from the application of the agreed upon search terms, the FDIC wanted to apply predictive coding to all 1.8 million documents, plus any additional ESI later collected by Progressive. The FDIC maintained that it would not have initially narrowed its search terms if it had known that Progressive would employ predictive coding because predictive coding achieves similar benefits as narrowed search terms. Additionally, the FDIC insisted upon incorporating the recommended practices of Equivio, Progressive's predictive coding technology vendor, into the protocol, which was lacking from Progressive's proposal. *Id.* at 4.

Therefore, the court had to determine whether Progressive could employ predictive coding and, if so, the contours of its usage. If the court concluded that Progressive could not employ predictive coding, then it would have to determine what Progressive's obligations were regarding these 565,000 documents.

The court held that Progressive could not employ predictive coding and, as contemplated by the ESI protocol, ordered Progressive to produce all 565,000 documents, responsive or not, to the FDIC. *Id.* at 11. The court reached this decision after noting that "[t]he cases which have approved technology assisted review of ESI have required an unprecedented degree of transparency and cooperation among counsel," which the court found lacking in this case. *Id.* at 4. The court acknowledged that there are many reasons litigators are unwilling to engage in this level of transparency and cooperation, such as:

[M]ethodological decisions reveal work product; discovery about discovery exceeds the scope of Rule 26 of the Federal Rules of Civil Procedure; revealing documents nonresponsive to discovery requests exposes

the producing party to unnecessary litigation risks; and the Federal Rules of Civil Procedure only require parties to conduct a reasonable search for responsive documents. *Id.* at 10.

Additionally, the court highlighted that Progressive's proposed plan for predictive coding was lacking due to its failure to incorporate the best practices recommended by Equivio, including Progressive's proposal to "give its counsel exclusive responsibility for training the predictive coding software." *Id.* The court also disapproved of Progressive's attempt to apply predictive coding after search terms had already been run. The court stated:

Progressive's proposal would relieve it of the burden of manual review of ESI according to the ESI protocol it stipulated to and allow it to adopt a predictive coding alternative to a small subset of the universe of ESI collected. *Id.*

The court also agreed with the FDIC that employing predictive coding would lead to numerous satellite disputes, such as whether a particular document is relevant or not. *Id.* at 11.

Regarding privilege, the court allowed Progressive to apply privilege filters to the documents and withhold documents these filters identified as "more likely privileged." *Id.* Progressive then had 90 days to compile a privilege log for these privileged documents, though there was a clawback agreement in place in the event of an inadvertent disclosure.

In light of the shortcomings of Progressive's predictive coding plan and in the absence of an agreement to modify the ESI protocol, the protocol as originally ordered maintained primacy. The court noted that "Progressive began utilizing predictive coding techniques to review ESI without the Defendants' agreement

continued on page 8

e-Pubs are coming to LJN!

Look for them online - coming soon!

Predictive Coding

continued from page 7

to amend the parties' stipulated ESI protocol Order (DKT # 67), and without seeking leave of the court to amend the ESI order." *Id.* at 2. Since, pursuant to this protocol, neither party had agreed to the use of predictive coding, the court was not inclined to incorporate predictive coding at this stage of the litigation over the FDIC's objections.

LESSONS LEARNED

Despite the court's rejection of predictive coding in this specific case, it did nonetheless look favorably upon predictive coding when employed in a manner more aligned

with recommended practices. The court stated: "Predictive coding has emerged as a far more accurate means of producing responsive ESI in discovery. Studies show it is far more accurate than human review or keyword searches which have their own limitations." *Id.* at 8. It went so far as to say that "[h]ad the parties worked with their e-discovery consultants and agreed at the onset of this case to a predictive coding based ESI protocol, the court would not hesitate to approve a transparent, mutually agreed upon ESI protocol." *Id.* at 9.

Therefore, litigants who think predictive coding may be needed in the course of discovery should be cooperative and transparent in the dis-

covery process. As demonstrated by Progressive, this includes working with the court and one's adversary regarding discovery methods and not unilaterally choosing a course of discovery, especially if doing so violates a mutually agreed-upon, court-ordered discovery protocol.

Ultimately, although it is an endorsement of predictive coding when used under the right circumstances, Progressive serves as a reminder that complications still exist when attempting to employ predictive coding, including debates over best practices, resulting satellite disputes, and the need to cooperate with one's adversary.



Equivio

continued from page 4

implementation of RVM's Structured Review approach, and in particular the use of Equivio's Relevance product for predictive coding, 61% of the "hot" documents actually found by the client's legal response team during the review process would have been missed because those documents contained *none* of the keyword search terms that were first used to focus the project. Those "hot" documents would have been missed by the defense had they not implemented the analytics-based strategy proposed by RVM.

Failure to identify such a large portion of responsive documents

using traditional methods would have been likely and would have put the client legal team at a significant and material tactical disadvantage to the government during witness interviews had the government found the documents a traditional search would have missed.

CONCLUSION

Responding to government investigations has traditionally required a costly and time-consuming review process with marked potential for human error. In this case, using a more traditional data search and review methodology, assuming that the resources were available, would have required a staff of 100 review attorneys and more than five months to complete the task.

In contrast, using Equivio analytics and predictive coding as part of RVM's Structured Review program, the legal team completed its first pass review in 17 days using only five attorneys. Employing these dramatic efficiencies in technology and time in the first pass review phase, the client was able to save nearly 50% of the total review costs that would have been incurred using a traditional approach. At the same time, compared with the results generated by the best possible keyword search used at the start of the project, RVM's approach identified more than twice as many critical documents used to prepare for fact witness depositions.



ESI

continued from page 6

communicants, most prevalent domains, date periods, etc. to remove content with similar characteristics from the universe of machine categorized non-relevant content. Removing this obviously relevant content through the use of an ECA tool can frequently produce a categorized universe that is much richer

in non-relevant content, allowing counsel to achieve statistical validation of categorization quality much faster than with traditional methods.

CONCLUSION

While the initial excitement surrounding ECA may have dampened, the benefits have only increased as the technologies have moved beyond version 1.0. In addition, the sophistication around not only the TAR offering, but also the SME's

driving the process has moved leaps and bounds from its initial entry into the market. We can learn more about our client's data and successfully navigate the ever-increasing data volumes by applying a more intelligent process workflow by not limiting ourselves to a single technology, but rather applying the benefits of both.



<p>To order this newsletter, call: 1-877-256-2472</p>	<p>On the Web at: www.ljnonline.com</p>
--	---