

# The Metropolitan Corporate Counsel®

National Edition

www.metrocorp counsel.com

Volume 22, No. 9

© 2014 The Metropolitan Corporate Counsel, Inc.

September 2014

## How To Review Five Million Documents In Three Weeks: A Case Study

Sanjay Manocha, Esq.

RVM

*This article was published as part of Equivio's "Predictive Coding Minus the Hype" educational series.*

### Introduction

Over the past ten years, the private sector has seen increasing activity from government departments and agencies exercising significantly broader regulatory and investigative powers over the private sector. It comes as no surprise that, with the advancement of information retrieval technologies, these departments and agencies are hiring sophisticated eDiscovery talent and deploying leading eDiscovery analytics technologies, such as predictive coding, natural language processing, conceptual clustering, near duplicate detection and e-mail threading, to analyze mountains

***The implementation of RVM's Structured Review program enabled five law firm associates, in less than three weeks, to examine a dataset of approximately five million documents.***

of data collected from companies in the course of government and regulatory audits and investigations, as well as civil litigation and criminal prosecution. The government is using these tools to better identify information that

may serve as evidence of wrongdoing or violations of federal laws. These technologies are allowing federal agencies to intelligently sort through much larger volumes of data in less time and at a significantly lower cost than ever before. As a result, companies are forced to deal with more frequent and arduous government requests.

Many companies subject to investigation are failing to implement the same analytics-based strategies employed by the government for one reason or another. For example, they might find

*Sanjay Manocha, Esq. is the Director of Discovery Analytics and Review at RVM Enterprises, Inc. He oversees the implementation of advanced analytics and predictive coding technologies in RVM's eDiscovery practice. His primary focus is the advisement and implementation of discovery analytics strategies for Am Law 100 firms and leading corporate legal departments.*



Sanjay  
Manocha, Esq.

themselves either unfamiliar with rapidly evolving technologies or unable to find and retain qualified technological and legal talent who understand how these approaches impact the fact discovery process. Oftentimes, stakeholders find themselves mired in discussions over whether such approaches are defensible and how such technologies actually work.

The reality is that the government is already up-to-speed and has a significant tactical advantage over defense counsel.

### The Challenge

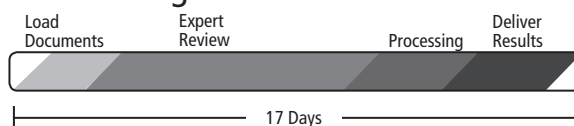
Rapid, thorough fact development and identification of the "hot" documents are always counsel's primary objectives. This goal can be put to the test when preparing for witness interviews in a Department of Justice investigation. Keywords are a good place to start. They can be easily applied, and almost immediately counsel can begin analyzing documents of potential interest. However, courts have cited studies demonstrating that, on average, keywords miss nearly 80 percent of the relevant documents in a collection.<sup>1</sup>

Additional measures to identify relevant documents are necessary for an effective defense. It is imperative that defense attorneys seek both efficiency and accuracy by exploiting state-of-the-art methodologies in order to appropriately prepare for such governmental agency investigations.

In a recent DOJ investigation, an Am Law 100 firm and one of the world's largest financial institutions partnered with RVM's Discovery Analytics team when they were presented with such a challenge.

With merely three weeks to review and analyze five million documents in preparation for substantive interviews requested by the Department of Justice, both the client and counsel realized that staffing a large attorney review team would not only have been difficult and exorbitantly expensive, but unlikely to be done in the necessary timeline. Counsel also expressed concerns that because the Department of Justice was likely using analytics, the defense would be at a tactical disadvantage if those analytics were to identify more relevant documents than a traditional approach.

### ►How long did all of this take?



**For more information, please email the author at [smanocha@rvminc.com](mailto:smanocha@rvminc.com) or Equivio at [info@equivio.com](mailto:info@equivio.com).**

To meet these challenges, RVM’s Discovery Analytics team was engaged to determine how best to proceed. After some initial testing and consultation, RVM recommended and counsel agreed to employ RVM’s Structured Review program (“RSR”). The following features were implemented for this particular challenge:

- Project-Level and Software-Level Iteration
- Keyword-Based Searches
- Equivio Near Duplicate and E-mail Thread Detection
- Equivio Relevance (Predictive Coding)
- Rules-Based Document Categorization and Truncation
- Conceptual Clustering
- Conceptual Categorization
- Conceptual Search

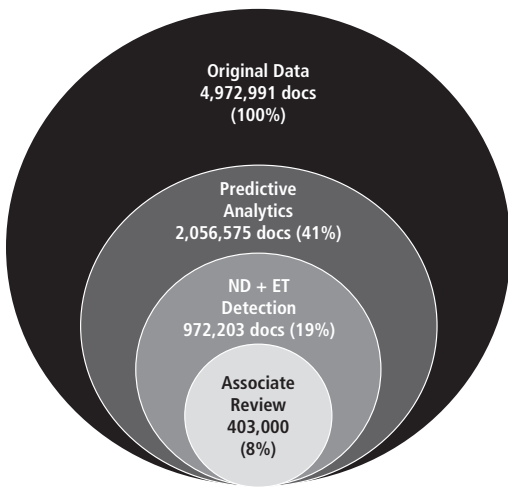
A key principle of RSR workflows is to eliminate text-level redundancies, topic-level redundancies and non-relevant topics using a portfolio of proprietary eDiscovery products and third-party technologies, such as Equivio Zoom. A commitment to these strategies results in a much smaller subset of unique documents across a diverse base of relevant topics. Once inessential documents are eliminated,

the case team can be trained on conceptual search analytics to help them quickly explore relevant topics and to tag documents as “hot” should they have a significant bearing on defense strategy.

Accuracy and defensibility are achieved by using each tool in an iterative fashion and refining its

own results and also by iteration at the project level, using results of one technology to improve the results of the other technologies employed.

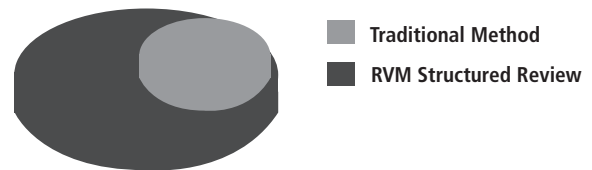
The implementation of RVM’s Structured Review program enabled five law firm associates, in less than three weeks, to examine a dataset of approximately five million documents. RSR reduced the population by 92 percent. Just over 400,000 documents remained that required attorney review. Further, analysis demonstrated that 61 percent of the “hot” documents found by counsel contained no keywords – in other words, these documents would have been missed by the defense had they not implemented an analytics-based strategy. Failure to identify such a large portion of the key evidence would have confirmed defense counsel’s fears – they would have been at a material tactical disadvantage to the government at witness interviews.



**The reality is that the government is already up-to-speed and has a significant tactical advantage over defense counsel.**

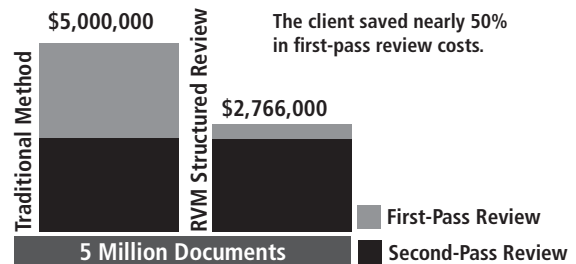
In sum, the firm discovered more material facts and “hot” documents than would have been possible using more traditional review methodologies and achieved substantial cost savings in the process while maximizing accuracy.

**61 Percent Of Hot Documents Would Have Been Missed If Not For RSR**



**The Resolution**

Responding to government investigations often requires a costly and time-consuming review process with marked potential for human error. Relying upon a more traditional methodology, assuming the resources were available, first-pass review for fact discovery would have taken a staff of 100 contract attorneys more than five months to complete. By employing RSR, the client was able to complete the first-pass review with a staff of 5 attorneys in 17 days saving nearly 50 percent of the total review costs that would have been spent using the traditional method, all the while identifying more than twice as many “hot” documents.



The client’s use of RVM’S Structured Review program resulted in a speedy review that saved costs, identified a large source of relevant documents that might not otherwise have been found and helped the legal team achieve a higher practice standard in this investigation.

1. See *Da Silva Moore v. Publicis Groupe*; No. 11 Civ. 1279 (ALC) (AJP), 2012 U.S. Dist. LEXIS 23350 at \*19 (S.D.N.Y. Feb. 24, 2012) (citing David L. Blair & M. E. Maron, “An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System,” 28 *Comm. ACM* 289 (1985), amongst numerous other sources).