A Look at How Workflow, Analytics, and CAL® Can Make a Difference in Cost Savings

How HaystackID™ Helped a Global Manufacturer Save Hundreds of Thousands of Dollars on a Mid-sized Litigation

By Adam Rubinger, JD, and Cameron Tschannen, Esq.

HAYSTACK™

A Mid-sized Matter That Could Have Cost Close to a Million Dollars

Snapshot Summary: Our client, a global manufacturer, approached HaystackID to assist with processing, early case assessment (ECA), hosting, and managed review for a mid-sized, time-sensitive litigation matter. Our client collected and securely transmitted approximately 650 GB of data, comprising several million documents, to HaystackID for processing and hosting. We then worked collaboratively to create workflow efficiencies that yielded considerable cost savings. Together, the team created a workflow that used both structured analytics and technology-assisted review (TAR) to create a substantial and documented cost savings in the high six-figures.

Collaborative Workflow Yielded Large Culling Results

Members of HaystackID's project management, review management, and analytics teams met with the client legal team and their outside counsel to discuss and plan a strategy with an eye towards cost savings. The first step in the process consisted of limiting the number of custodians that were collected. This targeted approach resulted in 650 GB sent to HaystackID, totaling approximately 4.3 million documents. Utilizing data reduction strategies at the time of processing including: de-duplication, de-NISTing, and date filtering, the data set was further reduced to 250GB comprising 1.65 million documents.

Working together to create a keyword search strategy that would further reduce the review corpus, the client, HaystackID, and outside counsel were able to further cull the data to a total review corpus of 540,000 documents. This represented an 88% reduction of the documents initially sent to HaystackID.

HaystackID then analyzed the proposed ESI Protocol of the case and discovered that the use of TAR was acceptable. HaystackID suggested the use of structured analytics, including email threading, to limit the review set to inclusive-only (unique) content. By removing the non-inclusive content from the review set, the result was a further reduction of 44,958 documents from review. The team also discussed and agreed to employ the use of Continuous Active Learning (CAL) to aid in the review of this matter.

As the review was set to begin, it was determined that certain search terms were overly broad and documents that hit solely on those terms could be excluded from review. This approach resulted in a further reduction of 204,028 documents from the review set.

The Impact of Continuous Active Learning

With a final review corpus of 253,786 documents amenable to CAL, the review phase of this litigation matter commenced. In addition to the CAL process, an alternate workflow was created and leveraged to streamline the review. This alternate workflow included file analysis performed to see if any documents could be mass tagged as not responsive. Exactly 9,273 documents were mass tagged as not responsive, with the remaining 3,037 documents being reviewed with document family members (if any family members were coded as responsive during the CAL workflow) or reviewed independently and sorted by textual near-duplicate grouping.

Once the CAL model was stable and the relevance rate of documents being reviewed via the model was sufficiently low to meet defensibility and proportionality standards—determined with the assistance of and at the direction of counsel—an elusion test was performed. Elusion tests measure the quality of the review effort by sampling the documents the CAL index deems non-responsive, which adds defensibility to the decision to conclude the review. The results of the elusion test were favorable as the elusion rate was just 0.13% and led to a decision to consider the review complete. Based on process, workflow, analysis, and consultation between HaystackID, the client, and outside counsel, 196,260 documents were determined not to require further review and were likely non-responsive. This equated to final total review of only 69,836 documents, resulting in a substantial savings in time and cost for the client.

Collaboration and CAL Lead to Significant Cost Savings

Estimating cost savings requires a combination of facts and assumptions in order to arrive at a reasonable estimate. One of the facts to consider was the overall review rate of 35 documents per hour. This number was based on the complexity of the coding panel and the number of issues to be coded for substantive review. When engaging in a CAL review, lower document per hour review rates are expected as they are a positive indicator that reviewers are primarily reviewing responsive documents and coding for complex issues. This is different than traditional review rates that do not include CAL, and many times, result in reviewers considering a much greater number of non-responsive documents. When reviewing a large number of non-responsive documents, the reviewers speed increases as non-responsive documents are easier to spot and code as such. The tradeoff of slower review rates for fewer documents is highly favorable in terms of time and cost savings to those matters not employing CAL.

Based on the total costs for the review and document per hour estimates, estimated cost savings can be derived. Taking the total number of documents after the first culling exercise (540,000 documents), and multiplying that number by the projected cost per document, allows for the determination of an estimated cost savings in this matter.

- Total docs in initial review corpus: 540,000
- Cost per document: $1.44
- **Total potential costs for review: $777,600**

- Total documents removed with mass tagging: 9,273
- Cost savings due to mass tagging: $13,353

- Total documents reviewed: 69,836
- Total review costs: $100,242
- **Total savings: $677,358**

As reflected through the lens of quantifiable facts and reasonable assumptions, the potential eDiscovery cost savings for this specific matter was approximately $677,000. This estimate assumes that after initial culling, use of search terms, and email threading, there was a potential review corpus of 540,000 documents.

As highlighted earlier, proactive planning quickly identified the salient opportunity to further cull the data by removing overly broad search terms. Based on this approach, the final document corpus for the employment of CAL was 253,786 documents. Additionally, after employment of CAL, the matter was able to reach conclusion after the review of only 69,836 documents. Provided below are the actual cost savings realized upon completion of the second culling effort. As shown below, the cost savings realized exceeded $265,000.

- Total docs in initial review corpus: 253,786
- Cost per document: $1.44
- **Total costs for review: $365,451**

- Total documents removed with mass tagging: 9,273
- Cost savings due to mass tagging: $13,353

- Total documents reviewed: 69,836
- Total review costs: $100,242
- **Total savings: $265,209**

From Planning Expectations to Repeatable Results

The proactive planning and practical protocol decisions on this project amplify how positive collaboration on workflow coupled with the use of CAL can provide significant cost savings in the execution of eDiscovery efforts. In this matter alone, the client was able to quantify a six-figure cost savings as opposed to potential costs using traditional, non-CAL enhanced review approaches. The practicality of the workflow, analytics, and CAL in this matter is not only repeatable, but to be expected when employed by our clients and HaystackID in mid-size and larger matters and where review sets are generally greater than 15,000 documents.

About the Authors

Adam Rubinger, JD is the Chief Client Experience Officer at HaystackID. Adam has over 20 years of experience and proven leadership in litigation support, information governance, as well as managing large-scale electronic discovery projects. Adam is a technologist, lawyer, and acknowledged expert in the areas of electronic discovery, advanced analytics, discovery workflow, and matter management.

Cameron Tschannen, Esq. serves as Director of Review at HaystackID. In this role he is responsible for managing the review of electronic data at the direction of, and in consultation with, the legal teams of clients and outside counsel. Cameron also helps expertly manage the application of analytics tools within eDiscovery platforms to ensure optimal review efficiency and quality control.

About HaystackID

HaystackID™ is a specialized eDiscovery services firm that helps corporations and law firms securely find, understand and learn from data when facing complex, data-intensive investigations and litigation. HaystackID mobilizes industry-leading computer forensics, eDiscovery, and attorney document review experts to serve more than 500 of the world's leading corporations and law firms in North America and Europe. Serving nearly half of the Fortune 100, HaystackID is an alternative legal services provider that combines expertise and technical excellence with a culture of white-glove customer service. The company was recently named a worldwide leader in eDiscovery services by IDC MarketScape and was included as a representative provider in Gartner's Market Guide for E-Discovery Solutions. For more information about its suite of services, including programs and solutions for unique legal enterprise needs, go to HaystackID.com.