

Blind De-Duplication

Unique Custom Scripting Avoids Critical Problems When Switching Vendors

Case matters can be in litigation for many months, if not years, which means that stability is one of the most important factors in choosing a litigation support vendor. Client firms place a great deal of trust in a vendor when they begin the processing for a new matter, and many issues can arise if a client needs to switch to another vendor for processing.

Moving from one vendor to another, or splitting a particular case between multiple vendors, can be problematic for many reasons. The most glaring one stems from de-duplication inconsistencies. That's because it's virtually impossible to use integrated tools to accurately test for duplicative records when any given file might be a duplicate of one processed with another vendor.

Background: *Relying on our expertise in de-duplication.* HaystackID has been asked to take over many projects in various stages of processing, so we've devised a defensible methodology for providing accurate de-duplication of records—regardless of whether a matter has been distributed to many different vendors. In one such situation, we were tasked with finding documents that were duplicates that we did not have in our system, had never seen before, and could not provide due to restrictive stipulations regarding the distribution of the documents.

HaystackID Responds: *Working with the most minimal amount of information.* Our HaystackID team asked the client for a list of the MD5 hashes—which are alphanumeric identifying keys that are unique to any given document—of the files in question. The MD5 hashes provided just enough information about the files to determine if the records were duplicative without needing the files themselves.

Using a stock HaystackID utility, the project management team set about converting the MD5 hashes into source data for a tag list. We avoided potential partial-family tagging by examining and then clearing all the tags in that family. We knew that if some docs within the family were not duplicates, the original documents in question were not from the same source as the family. Thus, the family should not be tagged.

Solution: *Separating duplicates from millions of documents.* The client was able to determine whether files existed in both data sets or only in one. This allowed them to perform an accurate de-duplication of records spread across multiple vendors. With our custom methodology and a common utility, HaystackID was able to examine a case involving several million documents for duplicates against tens of thousands of files that the project management team was unable to examine for themselves.

Outcome: *Lengthy, unnecessary processing times avoided.* Other vendors would have demanded access to the entire data set to perform an accurate de-duplication—or refused to take on the task. Thanks to our innovative and custom scripting, HaystackID enabled a common utility to eliminate excessive processing times to provide a comprehensive master list of unique documents in the record set. This saved our client precious time, needless aggravation, and unnecessary expense.

Learn More. Today.

[Contact HaystackID today](#) to learn more about how our highly rated eDiscovery consulting and services can enhance your ability to efficiently and economically achieve favorable investigation and litigation outcomes.

About HaystackID

HaystackID is a specialized eDiscovery services firm that helps corporations and law firms find, listen, and learn from data when they face complex, data-intensive investigations and litigation. With an earned reputation for mobilizing industry-leading computer forensics, eDiscovery, and attorney document review experts, HaystackID's Forensics First, Early Case Insight, and ReviewRight services accelerate and deliver quality outcomes at a fair and predictable price.

HaystackID serves more than 500 of the world's leading corporations and law firms from North American and European locations. Our combination of expertise and executional excellence, coupled with a culture of white glove customer service, makes us the alternative legal services provider that is big enough to matter but small enough to care. **Learn more today at [HaystackID.com](https://www.haystackid.com)**