



## 48,747 Discovery Documents + One Resource-Strapped Federal Agency = Uh Oh

### The Problem

IE Discovery had a client with a very challenging project. The client, a federal agency involved in a contract dispute with one of its suppliers, had just received its opponent's discovery production—48,747 documents in single-page TIFF image format, with document breaks but without digitized text. The client had only weeks to search the production for key evidence to use at an important hearing, and only four attorneys who could spend even a portion of their working day reviewing documents. With limited resources and a very tight budget, the client needed a strategy that would maximize the legal team's document review efficiency and best prepare them for their imminent hearing.

The legal team's initial review of the TIFF images uncovered wide-spread duplication within the discovery materials they had received. To speed the review and increase its efficiency, IE Discovery suggested that they use Equivio to identify and group related documents. Although unfamiliar with Equivio technology, the client agreed to give it a try.

### The Solution

Equivio's near-duplicate technology uses textual analysis to generate its results, so IE Discovery first ran the production's 200,000+ TIFF images through an OCR engine to generate searchable text. While the OCR process took approximately 24 hours to complete, subsequent Equivio processing took only a few hours. The entire Equivio analysis was loaded into IE Discovery's InfoDox™ review platform and was available to the review team less than thirty-six hours after the decision to implement Equivio had been made. Moreover, since the images had already been available in InfoDox™ for review, the



client was able to continue its review even as full-text search and near-duplicate groupings were loaded into the review platform.

Equivio’s near-duplicate analysis was valuable even for this relatively “noisy” document population. Traditional duplicate identification technology, such as MD5 file “fingerprinting,” didn’t work in this situation, as each TIFF image in the production contained a unique control number (i.e., a Bates number), meaning that no two image files could be exactly identical, even if they differed only by control number. In addition, due to differences in text formatting and source document condition, the OCR process itself introduced transcription errors, as the processing engines occasionally stumbled over unfamiliar names and terminology—or over-corrected for typographical errors within the original document. The resulting OCR text contained a great deal of extremely similar passages and documents, but no exact duplicates.

Given the limitations of OCR text and the duplicate-destroying control numbers on each page, IE Discovery chose a ninety percent (90%) similarity threshold to identify near-duplicate documents for purposes of creating Equivio Equiset groupings. Not every document fell into a near-duplicate grouping, but 13,787 of them (about 28% of the collection) did. To reduce the number of documents the review team needed to analyze, IE Discovery then used Equivio to identify the pivot document—the document containing the greatest amount of text—for each near-duplicate grouping.

For purposes of review, IE Discovery loaded only unique documents (i.e., those that did not fall within a near-duplicate grouping) and pivot documents into the InfoDox™ platform. If reviewers found that the pivot document was irrelevant to the dispute, they could eliminate entire clusters of documents from further review with a single click. Conversely, some groupings contained highly relevant materials, particularly groupings of contracts whose terms served as the basis for damages and whose terms and conditions changed over time. When a reviewer encountered one of these clusters via the pivot document, a single click in InfoDox™ would bring up the rest of the near-duplicate grouping for individual analysis.

### **Review Team Experience and Results**

Equivio near-duplicate analysis accelerated the review in two ways. First, focusing on ungrouped and pivot documents significantly reduced the number of documents that the client’s legal team needed to review in its first pass. Second, Equivio near-duplicate groupings also helped speed up the pace of the actual review. Within a grouping, for example, reviewers could see all iterations of the same document. After reading the pivot document carefully, subsequent near-duplicate documents could be reviewed quickly for differences, not laboriously reviewed from top to bottom as they would have been if they had come up as scattered documents. Neither IE Discovery nor the client tracked formal “before” and “after” metrics for the raw reviewer speed, but members of the review team



unanimously believed that they were able to moving through documents much more quickly because of the logical way in which they were able to sort and review materials.

Equivio analysis also helped the review team in another important, yet subjective area. Thanks to near-duplicate grouping, members of the review team felt significantly more confident in the consistency of their review. In addition, even when encountering a related but loose document that had fallen outside the 90% near-duplicate threshold, reviewers felt much more confident that they were making analytical calls that were consistent with similar documents they had previously seen inside a near-duplicate grouping.

Ultimately, the client believes that it saved hundreds of hours of review time by leveraging Equivio's near-duplicate analysis to increase document review speed. The client was able to complete its document review in a timely manner. An order hasn't yet been issued in the critical hearing, but the client feels certain that the document review they completed with help from Equivio, made a critical difference in their case development and hearing preparation. Indeed, based on this case, IE Discovery's client has since requested Equivio analysis for several additional cases.

IE Discovery, in turn, isn't surprised at the client's success. "We've never had a client who wasn't extremely satisfied with Equivio," says Stacy Jackson, Corporate Counsel for IE Discovery. "Once they've given it a try, they want to use Equivio in all their other projects."

## About Equivio

Equivio enables the management of data redundancy in content-centric business processes. Equivio's technology zooms in on unique data, allowing you to read less, think more, win big™.

With products for grouping near-duplicates, capturing email threads and automating prioritization, Equivio powers a broad range of business applications, including e-discovery, records management, email archiving, data retention and intelligence. To learn more about winning with Equivio, email [info@equivio.com](mailto:info@equivio.com) or visit [www.equivio.com](http://www.equivio.com).

Equivio Inc.  
5260G Nicholson Lane, Suite 150  
Kensington, MD 20895  
USA

[www.equivio.com](http://www.equivio.com)