



## Handling Near-duplicates at Munger, Tolles & Olson LLP

### Introduction

Visionary law firms recognize that they have a responsibility to their clients to actively promote the use of technologies that can reduce litigation costs. This is challenging, and in many cases unsettling, because new technologies require people to make changes. And people usually don't like change, even if the benefits are there. However, as this case study shows, near-duplicate technology has a compelling change-benefit ratio. As implemented at MTO, near-duping has minimal impact on the review environment and work flows, while significantly reducing the cost and risk of litigation.

### The Numbers are Staggering

All litigators are familiar with exact duplicates and de-duping. Everyone does de-duping of the review corpus to eliminate, or suppress, the exact duplicates. The technology has been around for years, it works well, and it's an accepted, standard part of the discovery flow. So what's a near-duplicate? Near-duplicates are files with small differences. The differences are usually in the textual content, although formatting differences also qualify. Near-duplicates turn up a lot in files like contracts, emails, order forms, proposals, document versions, and so on.



Intuitively, of course, people know about near-duplicates. Any litigation attorney who has spent those interminable late nights poring over endless documents in a case knows that déjà vu feeling that hits you around a coffee-inspired pre-dawn insight, “Haven’t I seen this document before, or something very similar...” The problem, of course, is much more severe when you have large review teams, of tens and even hundreds of lawyers. The near-duplicates are assigned randomly across all the teams, and there is no way to coordinate the review effort or, what is worse, to reconcile and integrate the information. But, until now, unable to quantify it, unable to fix it, the problem has been dealt with in an almost uniform manner throughout the industry – it’s been ignored. This is unfortunate because data from Equivio customers indicates that near-duplicates are likely the largest single addressable cost in litigation review.

### **Solution Exposes (and Solves) the Problem**

Always on the lookout for new technologies that can enhance the litigation process, the MTO litigation support team initiated the evaluation of Equivio software in the fall of 2006. Initial trials were positive. The first chance to take it live came through in late 2006, in a case for a major financial institution. The results were telling.

Ron Best, Director Legal Information Systems at MTO, “In our first case, a relatively small one, we had 100,000 documents. Using Equivio, we found that 28% of the documents were near-duplicates. This is indicative of how identification of near-duplicates can facilitate a more cost-effective litigation review processes.”

From the attorney’s point of view, “near-duping” reduces review costs by grouping the corpus into sets of very similar documents. We might have, for example, one group of 35 near-duplicates, all versions of a 20-page plan which had been circled around to various people in the organization for comments and corrections. In Equivio terminology, this is called an EquiSet. Once the EquiSets are in place, it’s a short path to reducing your review costs.

MTO’s team on this case was using Concordance as the review repository. Equivio is integrated with Concordance and all the major review platforms. For each document, there is a column which indicates the EquiSet to which the document belongs. This allows the reviewer to organize the documents by EquiSet. Similarly, documents can be assigned by EquiSet, so that a given reviewer will be assigned all the documents belonging to any given near-duplicate grouping. This helps organize the review process in a rational manner, ensuring a more coherent and more systematic review process.



The economics of near-duping don't stop there. Within each EquiSet, the reviewer is directed to the "pivot" document. The review of these near-duplicates starts with the pivot. The pivot is the most representative or most inclusive document in the set. The pivot structure allows a prioritized review. The pivot might be blatantly irrelevant. For example, one may encounter a personalized email invitation to an end-of-year party. Each email is slightly different, creating near-duplicates, because it contains the name of the addressee. Depending on the matter, it might be clear that this email has no bearing on the matter. Having read the pivot, the reviewer can make the call that this document is irrelevant, skip the rest, and move on to the next EquiSet.

### **Less Effort; Better Review**

Dealing with near-duplicates is very different from exact duplicates. With exact duplicates, it's essentially a case of deleting them. In some cases, the preferred treatment is suppression for review, and inflation for production. Near-duplicates are another proposition. The bad news is that each near-duplicate contains unique information, which might be decisive to the case. The good news is that Equivio actually reduces the risk of inadvertently missing this unique data.

Reverting to the example of an EquiSet with a 20-page plan: the reviewer starts by reading the pivot. If it is decided that the pivot is actually relevant to the matter at hand, then the reviewer can read each and every document in the EquiSet, as each one may contain unique information that is important to the case.

However, and here is the killer point, it's not necessary to read the options plan 35 times. First, the reviewer reads the pivot. From here on in, a compare tool can be used to highlight the differences in each document vis-à-vis the pivot. For example, a particular version may just have two words different. We simply zoom in directly on those two words. By the way, Equivio also supplies the compare tool, fully integrated, as part of the package.

Obviously, reviewing differences is a lot faster than reading complete documents. It takes seconds to review two words, half an hour to read 20 pages. The result is a much more efficient review process. No less important, it's also a lot more effective. As an attorney, one's worst nightmare is to miss those few words, tucked away deep in some document, which might swing the case against you. Equivio's near-duplicate groupings minimize this risk by allowing the attorney to pinpoint the unique words in each document.



### **Warning: Near-duping is Addictive**

In using the near-duplicate capability, Equivio received excellent feedback from MTO. In fact, as we have seen with MTO and other customers, once you've used near-duping, there's no turning back. It's actually a win-win technology for both law firms and clients. Clearly, clients are delighted by measures that help control litigation costs. Meeting tight time frames is also significant. With near-duplicate rates in the 20-30% range, near duplicate groupings can be crucial in completing the review on schedule.

From the client's point of view, another important factor is consistency of treatment. Ron Best added: "The near-duplicate groupings enable the consistent treatment of documents. In the normal course of review, the attorney will be tagging documents as responsive, privileged, etc. If, for example, the pivot document is deemed privileged, the Equivio groupings allow the reviewer to tag all the documents in that EquiSet as potentially privileged. This is an important consistency and quality assurance function." Near-duping is also the kind of technology that helps keep associates happy. Document review, the burden of which falls largely on the shoulders of associates, is not necessarily the most attractive proposition for budding litigation lawyers. Near-duping makes their lives easier, and, with risk reduced, less stressful.

People, both clients and staff, like the fact that near-duplicates are very intuitive. Near duping is not some vague or remote concept cooked up by engineers in a laboratory. On the contrary, near-duplicates are a problem from the real world that everyone understands. The ability to view the differences between documents makes it very concrete. There is also no "noise" in the Equivio near-duplicate sets because there are no false positives. As a result, from our experience, people feel confident about using the tool from day one.



## The Perfect Match: Technology Solves a Problem of its own Making

Technology is the medicine for a problem of its own making – the problem of near-duplicates. Digitization, email and networking have combined to generate the near-duping problem, a problem, albeit scarcely recognized, that is costing Corporate America dearly. Given the levels of near-duplicates, they represent probably the largest hidden cost factor in the entire discovery and review flow. With the emergence of Equivio’s technology, the industry is now in a position to eliminate this huge and unnecessary cost burden.

## 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)