

## To Cull a Mockingbird: Popular, but Risky, “Keyword” Collection Filter

Article By:

Ralph C. Losey

---

This is part four of the continuing series on two-filter document culling. Please read [part one](#) and [part two](#) and [part three](#) first. Hopefully you will like this *part four* sequel better than Harper Lee’s sequel.

### First Filter

Some first stage filtering takes place as part of the ESI collection process. The documents are preserved, but not collected, nor ingested into the review database. The most popular collection filter as of 2015 is still *keyword*, even though this is very risky in some cases and inappropriate in many. Quite frankly, keyword filtering and collection is an old black and white type version of e-discovery, one that is rarely used in large cases. Typically such keyword filtering is driven by vendor costs to avoid processing and hosting charges. I have a better solution to the cost issue, certainly better than the sequel to *To Kill a Mockingbird*, but elaboration would take us too far astray.

Some types of collection filtering are appropriate and necessary, for instance, in the case of custodian filters, where you broadly preserve the ESI of many custodians, just in case, but only collect and review a few of them. It is, however, often inappropriate to use keywords to filter out the collection of ESI from admittedly key custodians. This is a situation where an attorney determines that a custodian’s data needs to be reviewed for relevant evidence, but does not want to incur the expense to have all of their ESI ingested into the review database. For that reason they decide to only review data that contains certain keywords.

I am not a fan of keyword filtered collections, but in smaller cases it can still be appropriate. The obvious danger of keyword filtering is that important documents may not have the keywords. Since they will not even be placed in the review platform, you will never know that the relevant ESI was missed. You have no chance of finding them. See eg, William Webber’s analysis of the *Biomet* case where this kind of keyword filtering was used before predictive coding began. *What is the maximum recall in re Biomet?*, *Evaluating e-Discovery* (4/24/13). Webber shows that in *Biomet* this method First Filtered out over 40% of the relevant documents. This doomed the Second Filter predictive coding review to a maximum possible recall of 60%, even if it was perfect, meaning it would otherwise have attained 100% recall, which never happens. The *Biomet* case very clearly shows the dangers of over-reliance on keyword filtering.

Nevertheless, sometimes keyword collection may work, and may be appropriate. In some simple disputes, and with some data collections, obvious keywords may work just fine to unlock the truth. For instance, sometimes the use of names is an effective method to identify all, or almost all, documents that may be relevant. This is especially true in smaller and simpler cases. This method can, for instance, often work in employment cases, especially where unusual names are involved. It becomes an even more effective method when the keywords have been tested. I just love it, for instance, when the plaintiff's name is something like the famous [Mister Mxyzptlk](#). But, alas, that is very rare.

More in the next article on the dangers of keyword collection filtering, especially in complex cases when the keywords are untested.

Jackson Lewis P.C. © 2025

---

National Law Review, Volume V, Number 205

Source URL: <https://natlawreview.com/article/to-cull-mockingbird-popular-risky-keyword-collection-filter>