

EDD SHOWCASE: Highway Robbery?

Are we cheating clients if we don't purge duplicates

By Anne Kershaw & Joseph Howie

If you send an e-mail to three people, your system keeps one copy, with all three people listed in the "To" field. If you need to see all your e-mails to just one of those people, you search for that name. It's still just one copy. Indeed, it would be silly to have multiple copies. The same holds true for document discovery repositories, yet a recent survey suggests that lawyers still don't "get it," and often insist on keeping multiple, unnecessary copies.



While many techniques (e-mail threading, concept clustering, analytics) can improve review efficiencies, we focused on deduping across custodians, because it is readily available and can be implemented in most review programs. It is the low-hanging fruit of litigation cost control.

Lawyers who fail to check for duplicates across multiple custodians, instead removing only duplicates from within the records of individual custodians, end up reviewing at least 20% more records on average. Whether or not their document review bills are ever audited, these lawyers are not meeting their ethical obligations to both clients and the justice system.

In May, we surveyed electronic data discovery providers to learn what their clients requested with respect to management of duplicate records. Full results are available on the eDiscovery Institute website (www.ediscoveryinstitute.org). We received responses from ACT Litigation Services; Business Intelligence Associates; CaseCentral; Clearwell Systems; Daticon EED; Encore Discovery Solutions; Fios; FTI Consulting; Gallivan, Gallivan & O'Melia; Iris Data Services; Kroll Ontrack; Legal Document Management (LDM Global); Legal Document Services International; Rational Retention; Recommind; StoredIQ; Trilantic; and Valora Technologies.

A single example: If Bob sends an e-mail to three people, there are four total copies of that e-mail. When Bob backs up his copy, there are then five copies. Deduping only within individual custodians reduces Bob's two copies to one but there are still four copies to review, one for each custodian — what we would term quadruple billing. Deduping all custodians would result in just one file to review, a 75% reduction compared to single-custodian deduping.

The exact reduction in volume will vary with different collections. On average, single-custodian deduping removes one of five records; across-custodian deduping almost doubles that rate. Review costs are proportional to the volume reviewed; if a review based on single-custodian deduping costs \$500,000, deduping across custodians saves, on average, \$106,000, with prospects of saving up to \$200,000 or more.

Although all respondents offered across-custodian deduping, only 52% of projects received across-custodian deduping, 41% received single-custodian deduping, and 7% received no deduping. In light of the potential savings, the obvious question is, "Why?" Proponents of single-custodian deduping usually claim that it is necessary so lawyers can know which custodian had a copy of the record. Clearly, these folks have no understanding of databases and the fact that one field can hold the names of all the custodians who had a copy of the record (just like the "to" field in your e-mail). Lawyers can see all custodians — something that is very difficult to determine with a separate database record for each copy.

Type of Deduping	Average of Responses*		
	Average % Reduction	Minimum % Reduction	Maximum % Reduction
Single-Custodian	21.4	9.7	40.2
Across Multiple Custodians	38.1	22.6	62.7

*A full report on the survey is available online at ediscoveryinstitute.org.

Complaints that copies of the same file may contain minor differences in hidden metadata are easily addressed. First, metadata is almost never examined by reviewers doing normal relevance and privilege reviews. The occasional need to perform a forensic examination on some records is no reason to burden pre-production review with all possible forensic copies: parties can simply keep this metadata, and if it is needed, provide a database or delimited file containing just the metadata for each copy with fields for the custodian's name, the path and filename, and the record number of the deduped record produced.

Parties also can reserve rights, via agreement or court order, to request metadata pertaining to specific e-mails or files.

Another argument is that single-custodian deduping is required to satisfy court-imposed requirements to produce each custodian's copy. In the e-world, however, indicating all the custodians in a separate database or delimited file is producing each custodian's copy. If someone wants to print it three times, one for each custodian, so be it, but it's just silly to store it three times electronically.

We asked several judges to review this article and all quickly grasped the benefits of deduping across custodians. When asked if deduping practices should be considered when deciding attorneys fees, most indicated it would be appropriate.

Said U.S. Magistrate Judge John Facciola, "Certainly. I already look for ...over-lawyering, having too many people doing the same thing, or having overqualified people do what the more junior people should do. ...Failing to dedupe is the electronic version of the same problem."

The American Bar Association's Model Rules of Professional Conduct 1.5(a) states that "A lawyer shall not make an agreement for, charge, or collect an unreasonable fee or an unreasonable amount for expenses," and the first factor listed is "the time and labor required..." Considering that reviewers are examining exactly the same content multiple times, it is hard to argue that multiple reviews are necessary or ethical.

Rule 1.1 of the ABA MRPC requires that lawyers be competent. It is becoming increasingly clear that competency in the world of handling large electronic

collections includes being competent in selecting technologies and processes that are cost-effective.

Rule 1 of the Federal Rules of Civil Procedure states what should be the ultimate goal in dealing with EDD: to secure the just, speedy, and inexpensive determination of every action and proceeding.

Lawyers have slowly moved away from having to see a piece of paper for every record; now, they need to move away from having to see a unique database record for every copy of every e-mail or electronic document that is reviewed. And they need to move towards systems that permit multiple tables or views of the records they're charged with handling.

Anne Kershaw is principal of *A. Kershaw Attorneys and Consultants*, and co-founder of the *eDiscovery Institute*, both based in Tarrytown, N.Y. E-mail: Anne@AKershaw.com

Joseph Howie is the principal of *Howie Consulting*, based in Bartlesville, Okla. E-mail: Joe@HowieConsulting.com.

Law Technology News August 2009