

AccDocket

WWW.ACC.COM

THE JOURNAL OF THE ASSOCIATION OF CORPORATE COUNSEL

January/February 2010

INSIDE:
Canadian
Briefings

10 WAYS TO AVOID FRAUD OR CORRUPTION

Ethics and Ediscovery Review

10 Questions to Ask Before
Joining a Nonprofit Board

CLO Roundtable: Finding
Opportunities in the Midst
of Obstacles

250+ Canadian Things
GCs Need to Know

COOK

An illustration of a woman's head with red hair and a brown checkered beret, looking out from the top of a large green ceramic cooking pot. The pot has the word 'COOK' written on it in raised, orange-brown letters. The background shows a kitchen with a window, a desk with a computer keyboard, and a chair.



Much has been written about the staggering costs and burdens of electronic discovery. If you manage ediscovery for your company, someone, somewhere along the electronic discovery process line has likely raised the issue of “de-duplication” or “duplicate suppression.” If the decision

was made to “de-duplicate” only within the records of individual custodians — as opposed to across custodians — you are paying lawyers to review the same documents from multiple custodians, multiple times, thereby inflating costs and increasing the risks of inconsistent privilege, confidentiality and relevance decisions.



Reprinted with permission of the authors and the Association of Corporate Counsel: Patrick Oot, Joe Howie, and Anne Kershaw, “Ethics and Ediscovery Review,” *ACC Docket* Vol. 28, Issue 1 (Jan/Feb 2010): Pages 46-57. Copyright © 2010, the Association of Corporate Counsel. All rights reserved. If you are interested in joining ACC, please go to www.acc.com, call 202.293.4103 x360, or email membership@acc.com.

Ethics and Ediscovery *Review*

By Patrick Oot, Joe Howie and Anne Kershaw

A recent study¹ published by the Ediscovery Institute based on a survey of leading ediscovery providers (*Deduping Survey*) shows that, despite the technical ability to suppress or consolidate duplicates within an electronic document population, chances are about 50:50 that your outside counsel fails to take advantage of this technology, opting instead to double-

bill for reviewing unnecessary duplicates for privilege, confidentiality and relevance. The study shows that, on average, law firms that do not consolidate duplicates across custodians are reviewing 27 percent more records than needed, and in some cases 60 percent or more, raising serious ethical issues involving conflicts of interest and technical competency.



Background on Duplicate Consolidation

The bulk of this article involves the extent to which law firms and companies may be failing to consolidate duplicate electronic records before engaging in document reviews. The idea is pretty simple: If one of your employees emails four other employees, that email may be found in the records of those five employees during discovery. In the ediscovery review world, there is no need to keep and review all five copies. There is technology available today that is able to:

- identify the fact that all five records are virtually identical; and
- consolidate the information about all five copies in the repository database fields, so that when reviewers are looking at a record, they can see who had copies of it and in what folders or directories those people kept their copies.

If duplicate records are not consolidated, multiple reviewers will look at exactly the same content to make exactly the same decisions. Not only does this duplicate review efforts, it wastes time. This is not double-billing by analogy; it is double-billing by definition.

The *Deduping Survey* asked leading ediscovery vendors in May 2009 to provide metrics on the results they obtained using different treatment options for addressing duplicate electronic documents. The results provided data points on the reductions in the volume of ediscovery requiring review, depending on whether the duplicate consolidation was performed only within the records of single custodians or across all custodians.

To illustrate this point, assume that Bob is the author of the email referenced above. Because Bob backs up his hard drive, he has three copies of the email that was sent. Additionally, assume that the four recipients each have just the one copy they received. This would result in seven copies of the same email in the document system or with the custodian. Consolidating duplicates within custodians will result in only one copy of Bob's emails being reviewed when looking at Bob's documents, a savings of the review time for the other two copies. However,



PATRICK OOT is director of electronic discovery and senior counsel at Verizon in Arlington, Va. He is a co-founder of the Ediscovery Institute and a member of the advisory board for the Georgetown University Law Center CLE Program's Advanced Ediscovery Institute. He can be contacted at patrick@ediscoveryinstitute.org.



ANNE KERSHAW is founder of A. Kershaw, PC Attorneys & Consultants, a nationally recognized litigation management consulting firm, and is president and co-founder of the Ediscovery Institute, a not-for-profit corporation dedicated to testing the use of technology to resolve discovery issues. She also teaches at Columbia University and the Georgetown Ediscovery Academy. She can be contacted at anne@ediscoveryinstitute.org.



JOE HOWIE is founder of Howie Consulting, a company that identifies and promotes technology and processes for the cost-effective management of ediscovery, and is director of metrics development at the Ediscovery Institute. He can be contacted at joe@ediscoveryinstitute.org.

consolidating duplicates only within the records of single custodians will still leave Bob's copy as well as the four other copies stored by the recipients.

Below are the average results obtained in the *Deduping Survey*. (See Graph 1).

Each case will be different, and you cannot be sure what results will be obtained until the data is analyzed in a specific case; but the *average* volume reduction suggests that significant savings can be achieved by consolidating data across multiple custodians. For example, assuming that there were 100 Gigabytes of data before duplicate consolidation, failing to consolidate duplicates across custodians in this case would result in initial review costs that are about 27 percent higher *on average* than consolidating only within single custodians. (See Graph 2 on page 49). However, there are cases where the savings could be even more

compelling. Consider the table that shows the maximum reduction in volume achieved by specific vendors in specific cases as reported in the *Deduping Survey*: (See Graph 3 on page 49).

You can see that reductions in volume on the order of 60 to 70 percent or more are not uncommon, if a party consolidates duplicates across custodians.

Now the disturbing part of the survey: The respondents reported that duplicate consolidation across custodians was used in 51.9 percent of projects, with 40.6 percent receiving single-custodian consolidation and 7.6 percent receiving no duplicate consolidation. Assuming that billing for review is proportional to the volume of data being reviewed, it appears that in about half the cases the clients probably grossly overpaid for responsiveness and privilege reviews by about 27 percent, and they

Graph 1 Deduping

Type Deduping	Average % Reduction in Volume of Records from All Projects	Average of Minimum % Reduction in Volume of Records, All Respondents	Average of Maximum % Reduction in Volume of Records, All Respondents
Single-Custodian	21.4	9.7	40.2
Across Custodians	38.1	22.6	62.7

Graph 2 Consolidation

Beginning Volume before Consolidation	100 GBs
Volume after Reduction by Single Custodian Duplicate Consolidation (100 GB minus 21.4%)	78.6 GBs
Volume after Reduction by Across-Custodian Duplicate Consolidation (100 GB minus 38.1%)	61.9 GBs
Single Custodian as Percent of Across-Custodian (78.6 GB divided by 61.9 GB)	127 percent

might have lost the chance to cut their review bills by 50 or 60 percent or more.

This brings us to the potential ethical problems arising from the review of duplicate documents. Why is duplicate consolidation not being used in all cases or projects and across all custodians? The short answer is that some lawyers are not as technically competent regarding ediscovery and document review. This lack of technical knowledge is compounded by the fact that there are no financial incentives to become competent. After all, some would argue that this could result in lower billing for ediscovery in a time when other practice areas might not be generating as much revenue. As George Bernard Shaw said about surgeons:

That any sane nation, having observed that you could provide for the supply of bread by giving bakers a pecuniary interest in baking for you, should go on to give a surgeon a pecuniary interest in cutting off your leg, is enough to make one despair of political humanity. But that is precisely what we have done. And the more appalling the mutilation, the more the mutilator is paid. He who corrects the ingrowing toe-nail receives a few shillings; he who cuts your insides out receives hundreds of guineas...
—*The Doctor's Dilemma with Preface on Doctors (1910)*

In the case of ediscovery processing we have a setting where ediscovery vendors and law firms have a perverse incentive to be inefficient, especially if the corporation is busy minding other stores. It's a case of the half-empty/half-full glass. The *Deduping Survey* suggests that half the time duplicates are consolidated across custodians, regardless of this incentive, but that half the time it's not done.

The relevant ethical provisions taken from the ABA Model Rules of Professional Conduct are as follows:

Rule 1.1, Competence: "A lawyer shall provide competent representation to a client..."

If ediscovery were a small part of litigation and duplicate consolidation had an imperceptibly small impact on ediscovery, the whole debate might be dismissed under the rationale of *de minimis non curat lex*. However, the cost of ediscovery in general, and the cost of relevance and privilege reviews in particular, have been a major concern for years. There are no excuses for "not getting it" when it comes to ediscovery. Lawyers who bill hundreds of dollars an hour are implicitly promising a certain level of competence that would include the basic notion of consolidating duplicates. The survey shows that respondents had been offering duplicate consolidation for eight to 10 years unless they hadn't been in business that long. When lawyers and the courts were first grappling with electronically stored information (ESI), maybe there was an excuse for not having thought this issue through, but those days are gone.

Graph 3 Ediscovery Vendor Survey Results

Maximum reduction in ediscovery achieved by consolidating duplicates across custodians in specific cases.	
Vendor	Percent of Original Volume
Act	35
BIA	75
CaseCentral	56
Clearwell	65
Daticon	79
Encore	60
Fios	41
Gallivan, Gallivan and O'Melia	65
Iris Data Services	90
LDSI – no backup tapes	40
LDSI – with backup tapes	85
Recommind	60
StoredIQ	80
Trilantic	70
Valora	40

Rule 1.3, Diligence: "A lawyer shall act with reasonable diligence and promptness in representing a client."

Having multiple lawyers looking at exactly the same content, to repeatedly make exactly the same purpose, cannot be said to represent reasonable diligence and promptness. On average, responsiveness and privilege reviews could be taking 27 percent more time than necessary — and often quite a bit more. Every month a case is active, there are all sorts of costs not directly related to responsiveness and privilege review. Extending the review by 27 percent extends those costs as well.

If **duplicate records** are not **consolidated**, multiple reviewers **will look at** exactly the **same content** to make **exactly** the same decisions.

ACC Extras on...Ediscovery

ACC Docket

- *Corporate Strategies for Reducing Ediscovery Costs (Feb. 2008)*. This article discusses general ways costs associated with ediscovery can be minimized, including a records retention policy, a data map and establishing partnerships. www.acc.com/docket/corpstr_edisc_feb08
- *Recovering Ediscovery Costs (Jan. 2008)*. In-house counsel have amendments to the Federal Rules of Civil Procedure to contend with, and the high costs of complying with them tend to come from the legal department's pocket. Learn about how to recover from the cost of ediscovery. www.acc.com/docket/rec_edisc_jan08
- *Trends in Discovery of Electronically Stored Information (Jan. 2007)*. Get the Honorable David J. Waxse's perspective from the bench: What do lawyers overlook during ediscovery? www.acc.com/docket/trends_in_edisc_jan07
- *The Emerging Role of Office of Technology Counsel (May 2007)*. Corporations are facing challenges in litigation due to the changes in case law, as well as the amendments to the Federal Rules of Civil Procedure (FRCP). How should in-house counsel respond to this changing infrastructure that is expanding services and technologies offered by ediscovery providers? www.acc.com/docket/tech_office_may07
- *Rough Waters Ahead for Ediscovery and the New Federal Rules of Civil Procedure (Jan. 2007)*. The amendments to the pretrial discovery rules of the Federal Rules of Civil Procedure are supposed to clarify how to handle electronically stored information. However, the rules have stimulated a new breed of complex discovery dispute. www.acc.com/docket/newfed_rules_jan07

Program Materials

- *Ethics Issues Surrounding Electronic Discovery (Oct. 2008)*. We all know that compliance with the electronic discovery rules is a challenge. But what about the ethical issues we face as attorneys? How can we best preserve the attorney-client privilege? What are the ethical implications of outsourcing? How about our own technological competency

(or lack thereof)? This panel attempts to answer these questions and more. www.acc.com/pm/ethics_edisc_oct08

- *Hot Topics in Ediscovery: Are There Any Other Kinds? (Oct. 2008)*. Instead of shrinking in horror, let our panel of experts help you answer burning questions in the world of ediscovery including: Is that electronic evidence going to be authenticated and admissible in court? Isn't there a better way to find the data I'm looking for? www.acc.com/pm/hottopic_edisc_oct08
- *Getting the Information You Need: Ediscovery, Conducting Interviews, Handling Electronic Data & More (April 2007)*. Learn about specific tools and methods you can use to improve your fact-gathering skills and take home to share with your attorneys and/or business clients. www.acc.com/pm/info_u_need_apr07
- *Ediscovery MiniClinic: Your Roadmap to Success (Oct. 2007)*. A year after the new electronic discovery rules were enacted, many companies and their counsel began seeking answers to the ever-evolving problem of electronic discovery. This panel explores best practices in three segments of electronic discovery: securing documents and other information before it's too late, gathering what's needed when it must be gathered and analyzing information efficiently and effectively. www.acc.com/pm/edisc_mini_oct07

Webcast Transcript

- *The New Federal Ediscovery Rules: How Your Ediscovery and Erecords Management Practices Should be Revised for Compliance (Aug. 2006)*. This webcast discusses the practical effects of the FRCP amendments and how they impact existing practices concerning records management and ediscovery practices. www.acc.com/webtrans/edisc_compli_aug06

ACC has more material on this subject on our website. Visit www.acc.com, where you can browse our resources by practice area or use our search to find documents by keyword.



Click below to access each of the following Handbooks:

- Arbitration Handbook 2019/20
- Capital Markets Handbook 2019/20
- Competition Handbook 2019/20
- Construction and Projects Handbook 2019/20
- Corporate Real Estate Handbook 2019/20
- Data Protection Handbook 2019/20
- Directors' Duties and Liabilities Handbook 2019/20
- Dispute Resolution Handbook 2019/20
- Doing Business in... Handbook 2019/20
- Employee Share Plans Handbook 2019/20
- Employment Handbook 2019/20
- Finance Handbook 2019/20
- IP in Business Transactions Handbook 2019/20
- Labour and Employee Benefits Handbook 2019/20
- Leniency Handbook 2019/20
- Life Sciences Handbook 2019/20
- Media and Communications Handbook 2019/20
- Private Client Handbook 2019/20
- Private Equity Handbook 2019/20
- Restructuring and Insolvency Handbook 2019/20
- Structured Finance and Securitisation Handbook 2019/20
- Tax on Transactions Handbook 2019/20
- Venture Capital Handbook 2019/20

Free resources for ACC members at: www.acc.com/practicallaw



^{PLC}Cross-border Handbooks are multi-jurisdictional guides that draw on the editorial expertise of ^{PLC}Cross-border and the research experience of ^{PLC}Which lawyer?

Updated annually, each Handbook:

- Is dedicated to a particular legal topic or industry.
- Takes a comparative approach.
- Includes detailed Country Q&A analysis, Cross-border feature articles, analysis and profiles of leading practitioners (^{PLC}Which lawyer? and Lawyer profiles).
- Allows users to compare national regimes by question and jurisdiction through the ^{PLC}Country Q&A tool.

The Handbook series:			
Arbitration	Dispute Resolution	IP in Business Transactions	Private Client
Capital Markets	Doing Business in...	Labour and Employee Benefits	Private Equity
Competition Handbook	Employee Share Plans	Leniency	Restructuring and Insolvency
Construction and Projects	Environment	Life Sciences	Structured Finance and Securitisation
Corporate Real Estate	Finance	Media and Communications	Tax on Transactions
Data Protection	Insurance (coming soon)	Mergers and Acquisitions	Venture Capital
Directors' Duties and Liabilities	Investment Funds	Outsourcing	

Rule 1.5, Fees: (a) "A lawyer shall not make an agreement for, charge or collect an unreasonable fee or an unreasonable amount for expenses. The factors to be considered in determining the reasonableness of a fee include the following: (1) the time and labor required..."

Again, once a company has paid to have a record reviewed to determine its responsiveness or whether it's privileged or not, no further time is required, making it unreasonable per se to do it a second time or a third time or a fourth time. As will be discussed later, proper duplicate consolidation addresses the objections sometimes raised to performing duplicate consolidation across custodians.

Rule 1.6, Confidentiality of Information: (a) "A lawyer shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is permitted by paragraph (b)..."

In the context of ediscovery review, the gist of Rule 1.6 is that a law firm has implied authorization to disclose the client's records to reviewers to the extent those people are reasonably necessary to conduct the review. That consent should not be construed as authorizing disclosure to reviewers whose reviews are completely duplicative of other reviews. Real confidentiality is inversely proportional to the number of people who know the confidential information, and improper duplicate consolidation needlessly expands the circle of people who are given access to the confidential information.

Rule 3.2, Expediting Litigation: "A lawyer shall make reasonable efforts to expedite litigation consistent with the interests of the client."

This is a duty owed the tribunal and society in general. Lawyers are, after all, officers of the court. Performing duplicate consolidation only within single custodians does more than cost you money — it exposes you to all sorts of litigation risks. For example, the *TREC Legal Track Study* and other studies show that individual reviewers seldom agree with other reviewers more than 60 to 70 percent of the time. When multiple reviewers are examining virtually the same content, they are almost bound to make inconsistent decisions that increase the litigation risk to the company.

For example, one reviewer may decide a record is responsive, another one may decide a different copy is privileged in toto, a third may redact a copy of the document and a fourth may decide a different copy is nonresponsive. If the document was privileged, producing a copy may be deemed to have waived the privilege.

If it was responsive, producing one copy while withholding another could raise issues about the quality of the production and the motives behind not producing a copy of an admittedly responsive document.

One practice related to duplicate consolidation deserves special mention here: Sometimes parties will consolidate duplicates for purposes of their own review, but when producing to the other party they "redupe" or produce all of the individual copies of a document. They are, in essence, trying to bury their opponent with electronic records. To the extent that the requesting party has specified that it wants all the copies and the court has so ordered, the producing party has no practical choice. However, the lawyer's duty to expedite discovery ought to require the lawyer to at least raise the issue of duplicate consolidation with the other party and the court.

However, the lawyer's duty to expedite discovery ought to require the lawyer to at least raise the issue of duplicate consolidation with the other party and the court.

Rule 3.3, Candor Toward The Tribunal: (a) "A lawyer shall not knowingly: (1) make a false statement of fact or law to a tribunal..." and

Rule 4.1, Truthfulness in Statements to Others: "In the course of representing a client, a lawyer shall not knowingly: (a) make a false statement of material fact... to a third person..."

Under the new Federal Rules of Civil Procedure (FRCP) relating to ESI discovery, lawyers are making representations to their adversaries and to the courts regarding the volume of ESI that has to be handled and the time required to review those records. Lawyers who don't properly consolidate duplicates are inflating the time and cost required to review their productions.

Rule 1 of the FRCP states what should be the goal of lawyers not only in federal civil matters, but in the conduct of all their practice: "to secure the just, speedy and inexpensive determination of every action and proceeding."

Proper Duplicate Consolidation

We prefer the term "duplicate consolidation" rather

than the more common but less precise terms “deduping” or “deduplication” because deduping seems to have the connotation that reviewers will only be told about a single copy of a record. This connotation may have grown out of the limitations of some ediscovery processing software, which may have only been able to present information about one copy of a record. However, when duplicate consolidation is done properly, the reviewer is presented with information about *all* of the custodians who had copies of the record plus information about the folder or path where the record was stored.

In order to illustrate the point, let us take the following example of an email from one person to three others. In this case, Bob emails Vincent, Adam and Stevie, and those emails were all responsive to a discovery request. Proper duplicate consolidation would present information something like the following (the term “PST” indicates that the email was in the custodian’s Microsoft Outlook records):

Producing Custodians:

Bob, PST: Sent

Adam, PST: Inbox\Plant Explosion

Stevie, PST: Inbox\Potential Law Suits

This proper consolidation of information lets the reviewer see which custodians had copies of the record. In this case it makes it far more likely that the reviewer will note that Vincent never produced a copy of the email; that may in turn lead to the realization that the reviewer never sees anything produced from Vincent and may trigger an investigation into the reasons for this.

Seeing everyone who may have had a copy of a document can also be crucial when asserting privilege claims. In the above example, there could be a copy of an otherwise privileged email that is located in the records of a person whose possession of it could defeat the privilege. The company would be better served by knowing about such potential problems as early in the game as possible.

Proper consolidation lets a reviewer see all of the folders in which the duplicate document was placed. That contextual information is not readily apparent on the face of the document and can be very important. Done properly, duplicate consolidation permits reviewers to do a better qualitative job than without it — it’s not just about saving review dollars.

Proper duplicate consolidation should consider contextual information and any limitations of the software that the reviewers will be using to examine the documents. For example, lawyers typically want to make review decisions at the message attachment group level, i.e., looking at emails and their attachments as one logical unit, meaning that the email and attachments will all be treated alike, whether relevant or privileged. For documents that are attached to multiple emails, that may

mean that the review platform might contain a copy for each message attachment group to which the document belongs, depending on the flexibility and power of the review system.

The flexibility and power of the review platform and the preferences of the lawyers managing the project can also impact the treatment of emails for which there were blind or “BCC” copies. Some lawyers will prefer to have just the most complete copy of the BCC email, i.e., the one with the BCC recipients shown. Others will want to have both.

Regardless of exactly how these nuances are handled, the point is that you should consolidate duplicates to the maximum extent possible so as to lower the costs and time to review and to curtail additional litigation risks.

Sidebar 1: Sample Letter to Outside Counsel

To: Outside Counsel

Re: Controlling costs of processing ediscovery

Dear XXX,

We are reviewing strategies and processes that can help us manage the costs and risks of ediscovery. We want to ensure that our outside counsel are minimizing time spent reviewing multiple copies of the same records for privilege, confidentiality and responsiveness purposes. Accordingly, to ensure that effective methods are being used to properly reduce duplication in electronic records subject to discovery, please be advised that this consideration will be included in any future audit of your firm and our bills.

To the extent that you have already engaged in the analysis and elimination of duplication from electronic document reviews, we would appreciate a report of your findings, together with information as to how you currently consolidate information about the multiple copies of duplicate records that are processed. In addition, please advise as to your process for removing standard software-related files from document collections prior to incurring charges related to file size.

Finally, we believe we have an obligation to minimize the carbon footprint created by our litigation-related activities. In your report, please advise as to your practices with regard to printing paper copies of electronic-source records. We are specifically interested in whether you routinely make paper copies of all or most records that are reviewed for privilege or relevance.

Thanks, I look forward to hearing from you.

Sincerely,
In-house Counsel

Sidebar 2: Ediscovery Technical Competency Quiz

There is an ethical obligation to be competent in our representation of clients. If you are responsible for managing ediscovery as part of your corporate job responsibilities, you should be able to score at least 90 on this quiz. If your outside counsel control how your records are processed, they should also be able to score at least 90 on this quiz. (There are 11 questions worth 10 points each that a reasonably competent ediscovery practitioner should know and a more technical bonus question worth 20 points. See Sidebar 3 for answers.)

Select the single most correct answer (10 points for each correct answer):

1. One gigabyte of data

- a. May contain 25,000-500,000 pages, depending on the file types.
- b. Always contains at least 1,000 pages regardless of data types.
- c. Never contains more than 200,000 pages regardless of data types.

2. A *.exe file

- a. Can be safely ignored for ediscovery purposes.
- b. Could contain data.
- c. Should only be examined if found in the "My Documents" folder.

3. De-NISTing

- a. Refers to the use of data from the National Software Reference Library.
- b. Should only be done if there is a concern that an investigation may have criminal consequences.
- c. Is a seldom-used NASDAQ investigative procedure.

4. Deduplication

- a. Means producing the minimum number of copies of an item if there are duplicates.
- b. Means reviewing only the minimum number of copies of an item if there are duplicates.
- c. Means having record custodians only give you one copy of their files if they have duplicates.

5. Vertical deduplication

- a. Refers to consolidating duplicates within reporting relationships, e.g, within those custodians who report to the same person.
- b. Refers to consolidating duplicates from within the records of a single custodian.
- c. Refers to consolidating duplicates from within individual branches or folders of drives.

6. Hashing

- a. Refers to a process of scrambling a file for encryption purposes.
- b. Refers to the process of cracking a password on an encrypted file.
- c. Refers to a process of calculating a message ID for a file.

7. MD5 values

- a. Will always be different for different files.
- b. Ignore capitalization.
- c. Are highly unlikely to be duplicated if calculated from two different files.

8. SHA values

- a. Are more reliable than MD5.
- b. Are less reliable than MD5.
- c. Are on a par with MD5.

9. Deduping ESI across custodians for review

- a. Can, on average, lower review costs by 20 percent.
- b. Decreases risks of inconsistent review decisions.
- c. Lowers the number of people who are likely to have access to the records.
- d. Speeds review.
- e. None of the above.
- f. All of the above.

10. Digital forensics

- a. Typically involves copying a drive or device on a bit-by-bit basis.
- b. Is required in all high-dollar litigation.
- c. Requires the disclosure of the identity of the forensic examiner to opposing counsel.

11. Metadata

- a. Is critical in most non-criminal proceedings.
- b. Generally requires the services of a forensic examiner.
- c. Is sometimes used to refer to certain data within a file as well as data about a file maintained external to the file.

Bonus Question (20 point bonus): RFC2822

- a. Is an internet standard that can be used to consolidate emails that were produced from different email systems.
- b. Is an internet communications security standard used by the DOD and leading ESI providers for secure data hosting.
- c. Is an internet standard providing for secure VPN access.

Eliminating Software-Related Files Prior to Consolidating Duplicates

One step that should be performed before duplicate consolidation is to remove from consideration any files distributed with commercially available email software, e.g., Microsoft Word. The National Institute of Standards and Technology (NIST) maintains the National Software Library in conjunction with the US Department of Justice's National Institute of Justice. The NSRL contains "hash values"² for the files commonly distributed with many of the commercially available software packages. Such files can safely be eliminated from virtually all litigation as being irrelevant (unless of course the litigation pertains to one of the software packages and discovery requests call for files distributed by a specific software company). This process, commonly called "de-NISTing" should be the first step in cutting down the volume to be reviewed.

The Ethical Imperative to Conserve Resources

It is everyone's ethical obligation, not just lawyer,s, to conserve the earth's resources. However, despite the fact that most emails and other electronic business records were never printed out by the business people who created, sent and used them, there are large numbers of lawyers who have

all ediscovery records printed for review. This results in tremendous cost and potential delay to their clients. This is especially problematic when combined with other inefficient practices such as failing to de-NIST or consolidate duplicates.

Considering that one gigabyte of data will produce approximately 70,000 pages; and given that some cases run into the terabytes, it should be obvious that printing complete productions is wasteful and presents other risks, such as "Who controls the access to all that paper or how many more people does it take to use an inefficient review process?" Vendors who get paid by the record to process ediscovery virtually all work from monitors. This practice is faster with fewer risks of assigning the wrong codes or review decisions to the wrong records.

What Should You Do?

Assuming that you've been persuaded that you should to use duplicate consolidation across custodians for all your projects, what should you do? Here are some ideas:

- *Let outside counsel know that duplicate consolidation is vital to you.* Ask them what they're doing and how they're doing it. You will probably save review dollars just by raising the issue. A sample letter is shown in Sidebar 1.
- *Test the technical knowledge of anyone who is re-*

sponsible for processing or managing the processing of your ediscovery. A sample quiz is shown in Sidebar 2.

- *Measure and track metrics on ediscovery processing.* You should have a good feel for the volume reductions you can achieve in different types of cases with de-NISTing and different levels of duplicate consolidation. You should also know the metrics on review rates in terms of document decisions per hour for different types of cases. You are experienced enough to know that there will be good reasons for having different review rates in different cases, but at the same time metrics will be useful in developing projections and being able to recognize when projects are getting out of hand.
- *Review old cases.* You don't have to wait until you process new cases to develop a set of metrics for your company's records. Examine old cases to see what the level of duplicate consolidation across custodians would have been. Take a look to see how many duplicate documents received inconsistent treatment. You can confirm

Sidebar 3: Answers to Ediscovery Competency Quiz

1. a. Actual number of pages can vary widely depending on the file type. Some image files can contain many megabytes each. Simple text files may be very small, only 1 or 2 KB (i.e. about .000001 or .000002 GB) each.
 2. b. Users can save Word documents or other files with an ".exe" extension. To open them, they just change the extension back to the appropriate one, e.g. ".doc." The processing vendor should provide an exception report, listing all files with executable or unknown extensions, including their original file location.
 3. a. The National Institute of Standards & Technology (NIST) is the umbrella organization for the National Software Reference Library, which contains file names and hash values for files distributed with many software packages.
 4. b. Parties can dedupe for review purposes but still produce multiple copies of each selected record.
 5. b.
 6. c. For example, the acronym "MD5" was formed from the words "Message Digest."
 7. c. It is theoretically possible for two different files to have the same hash value, but it is highly unlikely.
 8. a.
 9. f.
 10. a.
 11. c.
- Bonus: a.

Consolidating duplicates across all the parties could **lead to smaller data** collections that **can be analyzed** much **easier** than the older approach of having **separate databases** for each party.

for yourself if the concerns raised in this article are speculative or real. You may want to ask for an adjustment on prior legal review bills based on what you find. At the very least you may want the information available when selecting future outside counsel. If your review system will not let you generate these metrics quickly and easily, you should consider changing your system or process. As management guru Peter Drucker said, "You can't manage what you don't measure." Under that rationale, without metrics you're not really managing the process.


- *Consider alternative billing arrangements.* As long as you pay for legal services by the hour, legal service providers will want to sell the most hours they can. Consider other arrangements such as per-record review fees based on deduped or consolidated records or a lump sum for ediscovery. Having metrics on other reviews as suggested above will help you assess what reasonable fees should be.
- *Consider consolidating duplicates across parties.* If you have litigation where the business people were communicating with each other before litigation, there can be an appreciable subset of records that will be produced by both or multiple parties. Consolidating duplicates across all the parties could lead to smaller data collections that can be analyzed much easier than the older approach of having separate databases for each party. This could be done in conjunction with either a shared repository or a shared vendor.

Maybe More Technical Stuff than You Wanted to Know

So far we have kept the discussion at a fairly high level. However, there is one topic that could save your company significant review dollars, but which necessarily involves a bit of a "deeper dive." It relates to the situation where your company or its employees may have multiple email systems in use, e.g., one department or division may use Lotus

Notes while another may have Outlook servers or different sections may have different versions of Outlook. Since “hash values” are so sensitive to even minor changes in a file (e.g., if a single quote is replaced by a double quote, that will cause the two files to have different hash values), the same email stored in different email systems may yield different hash values. Considering that email is the predominant type of electronic discovery, this can have serious impact on how many records you review.

The answer to this dilemma is to convert the copies of the emails to the form they would have been while in transit from the sender to the recipients. There is an internet specification called RFC 2822 that details how all the information pertaining to an email should be “packaged” for transmittal. Different email systems may store different parts of the email differently, but emails should be packaged the same way for transmittal, and by converting emails to RFC 2822 format, much higher levels of duplicate consolidation can be achieved. For example, an article by John Martin indicated that on one small project involving just three custodians, single-custodian consolidation achieved 4.5 percent reduction in volume, standard across multiple custodians achieved a total of

11.7 percent, but when the emails were converted to RFC 2822 format, the consolidation level was 47.2 percent. 

Have a comment on this article? Email editorinchief@acc.com.

NOTES

- 1 *Report on Kershaw-Howie Survey of E-Discovery Providers Pertaining to Deduping Strategies*. The study report published by the Ediscovery Institute was written by authors Joe Howie and Anne Kershaw, who also conducted the study. Results were received from ACT Litigation Services, Business Intelligence Associates (BIA), 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 (LDSI), Rational Retention, Recommind, StoredIQ, Trilantic and Valora Technologies.
- 2 “Hash values” are electronic “fingerprints” of files. While it is theoretically possible for two different files to have the same hash value, it is considered a virtual impossibility. Using hash values to identify and consolidate duplicate records is completely defensible whether you use the older MD5 (message digest) algorithm that results in 128-bit hash or the newer SHA (secure hash algorithm) that can produce longer and hence more secure hashes.