

Report on Kershaw-Howie Survey of E-Discovery Providers Pertaining to Deduping Strategies

By

Anne Kershaw

anne.kershaw@akershaw.com

www.akershaw.com

and

Joe Howie

joe@howieconsulting.com

www.howieconsulting.com

Copyright Electronic Discovery Institute 2009. Used with permission, other uses prohibited without the express written permission of the Electronic Discovery Institute. www.ediscoveryinstitute.org

Contents

Contents	2
Overview of Results	3
Survey Instructions	6
Respondents	7
Actual Survey	8
Initial Survey Results, Questions 7 and 8, Time Offering Deduping Services	10
Responses to Questions 9A, 9B and 9C, Reduction Levels Achieved	11
Response to Question 10, Percentage of Projects Receiving Different Deduping Approaches	12
Responses to Question 11, Why don't attorneys elect to use across-custodian deduping on all projects?.....	13
Response to Question 12, Do you provide reduping services?	16
Response to Question 13, Any other thoughts or insights?	17

Overview of Results

In May, 2009 Anne Kershaw and Joe Howie surveyed leading providers of e-discovery services and software to learn what their experiences had been in reducing the volume of ESI to be reviewed by deduping across custodian. They received useful metrics from 15 providers. All of the respondents provided deduping within the records of individual custodians and deduping across multiple custodians. They reported that on average volume was reduced by 21.4% by deduping within single custodians and that on average, deduping across all custodians achieved a 38.1% reduction, almost doubling the reduction achieved by single-custodian deduping.

In rare cases deduping across multiple custodians did not improve the reduction in volume achieved by just deduping within individual custodians. However, most respondents reported that the most reduction they had observed in projects with deduping across multiple custodians was in the range of 60-80% -- a truly phenomenal savings.

Considering that once the process is in place it doesn't cost anything to attempt deduping across custodians and that on average it reduces volume by 38% and that it could reduce volume by two-thirds or more, it is puzzling why it is not always done.

The average percentage of projects receiving single-custodian deduping was 40.6% with 51.9% receiving deduping across custodians, and 7.1% receiving no deduping. Interestingly, a respondent from the UK indicated that almost all UK cases were deduped across custodians but only 25% of the cases processed across custodians for the US.

All but one of the respondents offered "re-duping" – being able to produce a copy of a document that had been deduped across custodians as if the review had used custodian-level deduping. This basically differentiates between duplicate consolidation for review purposes vs. deduping for production purposes.

Reasons to dedupe across custodians:

- **Consistency.** There are litigation risks inherent in not treating identical records the same way, e.g. privilege waiver.
- **Comprehension.** It is easier to see who all had a specific record if there is one record with a single field containing all of the names of the custodians who had a copy of a document. Otherwise reviewers have to look at multiple records and they will miss certain patterns, e.g. the custodian for whom no records were produced.
- **Costs.** There are typically costs, direct or indirect of having data collections that are larger than necessary, e.g. storage, conversion to tif, inefficiencies in reviewing non-production query results.
- **Confidentiality.** The smaller the population to review, the fewer people need to see any of the records. More people means less more risks that confidential information becomes known.
- **Condition for Further Efficiencies.** Emails are the single largest record type in e-discovery. Presenting emails in the context of threads is essential to rapid and accurate review of emails, but unless duplicates have been consolidated the thread depiction is made virtually impossible.

Addressing reasons cited for not deduping across all custodians for all projects

Here are reasons given for single-custodian deduping and our response to the reasons.

- **Wanting to see who had a copy of a document:** This can be achieved by having a multiple-value field with the names of all custodians who had a copy. In fact this would provide more useful information than requiring reviewers to examine each custodian's copy of a record to see who all had a copy.
- **Wanting to see documents in the context of attachment relationships.** Deduping across custodians could take attachment relationships into account when deduping. A record could be included each time it was in a different attachment context. For example a Word document that was attached to two different emails and was in the My Documents folder could be presented in three times, once with each email and once as a free-standing or loose document. This would still permit substantial reductions compared to single-custodian deduping.
- **Complying with obligations to produce by custodian.** Treatment for review purposes does not have to be the same as treatment for production purposes, the producing party can always "re-dupe" if the requesting party insists on being inefficient.

Terminology

We use the term "single-custodian" deduping to mean deduping within a single custodian only. This is sometimes called "vertical" deduping.

We use the term "deduping across custodian" to mean deduping that examines the records of all custodians and is not confined to an examination of just one custodian. This is sometimes called "horizontal" or "project" or "corpus-wide" deduping.

Data Consistency

If we repeated the survey we would clarify instructions on how we wanted respondents to calculate the average gain from deduping across custodians vs. just within single custodians (question 9C). In some cases the values that respondents reported for deduping across custodians compared to just within single custodians was more than the difference between the averages for across custodian deduping less the average for deduping within single custodians. The reason was that they looked at actual cases where both single custodian and across custodian deduping was performed to calculate this value. Obviously there were cases where single custodian deduping didn't yield a large reduction but the across custodian deduping approach did.

Data Consolidation/Suppression/Views

The term deduping implies that only one record from a number of duplicates is placed in a database. However, if you read the responses from the respondents, you can see that some of them use advanced functionality such as placing all instances in the database but in essence tagging all duplicates with the tags associated with any of the instances, or of keeping all instances but only displaying one copy for certain purposes. We would encourage you to read all the narrative comments provided by the respondents as they contain several good insights.

Spelling

We found it annoying to hyphenate deduping and hence didn't.

Recommended Reading:

The narrative comments made in response to questions 11 (why attorneys don't elect across-custodian deduping) and 13 (other thoughts or insights) are well worth reading and studying particularly from providers who use advanced processing and presentation technologies.

Survey Instructions

The survey was emailed to respondents with the following instructions:

Subject: PARTICIPATION REQUESTED: LAW TECHNOLOGY NEWS SURVEY

SENT ON BEHALF OF ANNE KERSHAW AND JOE HOWIE

Dear _____,

We're preparing an article on cost-effective e-discovery review technologies. The article will be submitted to Law Technology News for its August EDD Showcase. As part of our preparations, we're gathering information from leading EDD providers on technologies or processes used to reduce the volume of data to be reviewed prior to production, with an emphasis on different types of de-duping. We'd appreciate it if you could fill in this form and click on the "submit" button in the upper right hand corner to submit it - it will be emailed to Joe Howie. We would appreciate it if you could do this by May 8, 2009. Please indicate if you'd care to be interviewed for and/or quoted in the article.

Thank you in advance for your participation.

Regards,

Anne Kershaw/A. Kershaw P.C.
Joe Howie/Howie Consulting

About Anne Kershaw

Anne is an attorney and senior consultant at A. Kershaw, PC//Attorneys & Consultants and the a co-founder of the EDiscoveryInstitute.org, a non-profit organization which researches the effectiveness of various litigation review technologies and processes. She teaches at Columbia and Georgetown Universities, is an expert on electronic discovery issues, and she consults with corporations and law firms on cost effective litigation management and discovery processes.

About Joe Howie

Joe is a technology consultant and writer who serves on the Editorial Board of Law Technology News. He is also a co-founder of the Association of Litigation Support Professionals. He edits and contributes articles to Update, the monthly electronic newsletter of the ALSP. His articles have been published by the ABA, CCA, AAJ, DRI and others.

Respondents

Company	Representative	Website
1. Act	Craig Lee, Director	www.actlit.com
2. BIA	Erin K. Smith	www.biaprotect.com
3. CaseCentral	Steven D'Alencon, Chief Marketing officer	www.casecentral.com
4. Clearwell	Kurt Leafstrand, Director, Product Management	www.clearwellsystems.com
5. Daticon	Ken Sokol, Director, Product Management	www.daticon-eed.com
6. Encore	Hemanthkumar Salem, Vice President, Professional Services	www.encorediscovery.com
7. Fios	Michael Harnish, Chief Technology Officer	www.fiosinc.com
8. FTI Consulting	Adam Cohen, Senior Managing Director	www.fticonsulting.com
9. GGO	Bill Gallivan, CEO	www.ggollc.com
10. Iris Data Services	Major Baisden, President	www.irisds.com
11. Kroll Ontrack	Michele Lange, Esq., Director of Legal Technologies	www.krollontrack.com
12. LDM	Paul L. Renehan Jr., International Director of Professional Services	www.ldmglobal.com
13. LDSI	Deborah Coram, Managing Director	www.ldsi.com
14. Rational Retention	Michael McCreary, Chief Operating Officer	www.rationalretention.com
15. Recommind	Craig Carpenter, VP and General Counsel	www.recommind.com
16. StoredIQ	Ursula Talley, VP Marketing	www.storediq.com
17. Trilantic	Nigel Murray, Managing Director	www.trilantic.co.uk
18. Valora	Sandy Serkes, President & CEO	www.valoratech.com

Actual Survey

The actual survey was sent in PDF form based on the following:

Cost-Effective E-Discovery Review Survey

We're preparing an article on cost-effective e-discovery review technologies. The article will be submitted to *Law Technology News* for its August EDD Showcase. We're gathering information for the article from leading EDD providers on technologies or processes used to reduce the volume of data to be reviewed prior to production, with an emphasis on different types of de-duping. We'd appreciate it if you could fill in this form and click on the "submit" button in the upper right hand corner to submit it – it will be emailed to Joe Howie. We would appreciate it if you could do this by May X, 2009. *Please indicate if you'd care to be interviewed for and/or quoted in the article.*

Definitions:

- **De-duping for review:** process in which duplicate records are not presented multiple times during the privilege or responsive reviews prior to production.
- **Custodian de-duping:** de-duping process in which only the records of a single custodian are analyzed to determine if they may be duplicates of other records. Sometimes also referred to as "vertical" de-duping. If multiple custodians have the same record, it would be presented at least once per possessing custodian.
- **Project de-duping:** de-duping process in which the records of multiple custodians are analyzed to determine if they are duplicates of other records. Sometimes also called "horizontal" de-duping. The same record is not presented to reviewers additional times just because it was held by multiple custodians.
- **Re-duping.** Process of producing to opposing parties multiple instances of de-duped records that were selected as responsive.
- **You.** "You" refers to the organization of the survey respondent.

Instructions: Enter Values, Click on "Submit" in the upper right hand of the PDF.

Survey Questions

Survey Respondent Information		
1	Name:	
2	Title:	
3	Organization:	
4	Phone:	
5	Email:	
6	Would you care to be interviewed and quoted for this article?	Yes No

Survey Questions		
7	Do you offer custodian de-duping for review purposes?	Yes Since: No
8	Do you offer project de-duping for review purposes?	Yes Since: No
[Continued Next Page]		

9	Assuming you have a population of records that are responsive but have not been de-duped, in your experience, what percentage reduction in volume is achieved by:	[Enter Responses Below]
9A	<ul style="list-style-type: none"> Just Custodian De-Duping compared to no de-duping: 	Average % reduction: Minimum % reduction: Maximum % reduction:
9B	<ul style="list-style-type: none"> Project de-duping compared to no de-duping: 	Average % reduction: Minimum % reduction: Maximum % reduction:
9C	<ul style="list-style-type: none"> Project-level de-duping compared to custodian-level de-duping: 	Average % additional reduction: Minimum % additional reduction: Maximum % additional reduction:
10	What percentage of your projects receive custodian de-duping vs. project-level de-duping?	Custodian: Project: No de-duping:
11	Why don't attorneys elect to use project-level de-duping on all projects?	
12	Do you provide re-duping services?	Yes Since No
13	Any other thoughts or insights you'd care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:	

Thank You! We will notify you when the article is published.

Initial Survey Results, Questions 7 & 8, Length Offering Deduping

The initial results are summarized below:

Company	Q. 7. Have Offered within-custodian deduping since	Q. 8. Have offered Across-Custodian Deduping Since
1. Act	2000	2000
2. BIA	2002	2002
3. CaseCentral	2004	2004
4. Clearwell	2006	2006
5. Daticon	1999	1999
6. Encore	2003	2003
7. Fios	2001	2001
8. FTI Consulting	2001	2001
9. GGO	2002	2002
10. Iris Data Services	Since inception	Since inception
11. Kroll Ontrack	1999	1999
12. LDM	1999	1999
13. LDSI	2003	2004
14. Rational Retention	Yes, no date provided	Yes, no date provided
15. Recommind	2007	2007
16. StoredIQ	2006	2006
17. Trilantic	2005	2005
18. Valora	2004	2004

Responses to Questions 9A, 9B and 9C, Reduction Levels Achieved

Company	Reductions Achieved by De-Duping as a Percentage of Number of Original Documents in Culled Data Collections								
	Q. 9A Single-Custodian Deduping			Q. 9B Across-Custodian Deduping			Q. 9C Across-Custodian Compared to Single-Custodian		
	Avg.	Min.	Max.	Avg.	Min.	Max	Avg.	Min.	Max.
1. Act	15	10	20	25	20	35	10	10	15
2. BIA	25	10	50	40	15	75	15	5	25
3. CaseCentral	19	3	21	51	44	56	32	41	36
4. Clearwell	25	15	35	55	45	65	30	30	30
5. Daticon	36.8	0	74	46.8	0	79	10	0	5
6. Encore	20	10	50	25	10	60	20	10	50
7. Fios	12	4	23	30	24	41	18	20	18
8. FTI Consulting*1									
9. GGO	20	10	50	40	10	65	20	0	15
10. Iris Data Services	13	4	58	33	20	90	23	16	76
11. Kroll Ontrack*2									
12. LDM Global	20			35			15		
13. LDSI – no b/u tapes	1-2		10-15	1-15	2-25	40	6.5		27.5
LDSI – w/b/u tapes	60	50	75	70	60	85	10	10	10
14. Rational Retention*3									
15. Recommind	5	0	20	30	10	60	25	10	40
16. StoredIQ	35	10	60	50	20	80	15	10	20
17. Trilantic	20	5	30	40	5	70	20	0	40
18. Valora	15	5	25	30	20	40	17.5	5	30
Totals	342	136	603	608	316	941	287	167	437
Number Responses	16	14	15	16	14	15	16	14	15
Average of Responses	21.4	9.7	40.2	38.1	22.6	62.7	17.9	11.9	29.2

Note: Yellow background indicates value was calculated from 9A and 9B as survey respondent did not provide the values for 9C

Example:

Footnotes

*1. Responded with See #13.

*3. Left blank. See comments.

*2. Responded with "Varies"

Responses to Question 10 about Percentage of Projects Receiving Different Deduping Approaches

Company	Percentage of Projects Receiving:		
	Single-Custodian Deduping	Across Custodian Deduping	No Deduping
1. Act	45	45	10
2. BIA	20	70*	5
3. CaseCentral	60	40	0
4. Clearwell	0	100	0
5. Daticon	56	13	31
6. Encore	49	49	2
7. Fios	70	20	10
8. FTI Consulting *1			
9. GGO	90	10	1
10. Iris Data Services	20	75	5
11. Kroll Ontrack	90	8	2
12. LDM Global	25	75	
13. LDSI - UK	1	98	1
LDSI - US	75	20	
14. Rational Retention *2	-	-	-
15. Recommind	15	80	5
16. StoredIQ	60	15	25
17. Trilantic	10	90	
18. Valora	5	70	25
Total	691	883	121
Number Responses	17	17	17
Average	40.6	51.9	7.1

Footnotes

* Calculated. Had been left blank but response to Q. 11 indicated that most customers deduped across custodian so we subtracted 20 (single custodian dedupe) and 5 (no dedupe) from 100.

*1. Indicated "See Response to #13.

*2. See response to #13.

Responses to Question 11, Why don't attorneys elect to use project-level deduping on all projects?

Company	Response to "Why don't attorneys elect to use project-level deduping on all projects?"
1. Act	Depending upon the project needs; (cost and defensibility) and the intended use of the review database and also production requirements.
2. BIA	Most of our clients choose to de-dupe across custodians - to ensure consistency in review tagging and redaction.
3. CaseCentral	Risk aversion - attorneys would often rather see all documents in context of custodian, especially if they need to produce by custodian or are reviewing key custodial data in preparation of depositions. (Custodial productions are often required by government agencies, in our experience.)
4. Clearwell	In the case of Clearwell, customers always elect to use project de-duping. That's because we ELIMINATE THE RISK of doing it by tracking all occurrences (with sources) of the original document. Then users can re-dupe by custodian on export.
5. Daticon	<p>The major benefit of project de-duplication is the elimination of redundant review. However, the process of project de-duplication has a downside when you consider that many document productions in litigation are organized by custodian. With project de-duplication, only one custodian's version of the document is physically present in the review process. Therefore, to facilitate a custodial document production, re-duplication must be applied to re-populate custodial versions of each document. Re-duplication is a complex process, which can result in errors and add time to the production cycle.</p> <p>Daticon EED's Discovery Partner offers the unique capability of managing groups of duplicate documents as a single unit. As a result, when using custodian de-duplication (or no de-duplication at all) reviewers need only review a single representative document from each duplicate group – achieving the same primary benefit delivered by project de-duplication. Furthermore, the use of custodial de-duplication eliminates the need for complex, potentially error-prone and time consuming re-duplication processes at the time of production. With the capability of managing duplicate document groups provided by the Discovery Partner platform, custodial de-duplication has become a nearly universal practice for Daticon EED clients. While the company does have the capability to provide project de-duplication services, its clients rarely choose that method.</p>

Company	Response to “Why don’t attorneys elect to use project-level deduping on all projects?”
6. Encore	<p>- The need to review documents within the context of a custodian's documents.</p> <p>- When the custodian list for a matter is evolving</p>
7. Fios	<p>In most cases, the production orders require the electronic evidence to be produced for specific custodians. If attorneys requested a de-dupe across the project, critical evidence for the required custodians could be omitted from the production set and therefore be in violation of the court order or discovery agreement.</p>
8. FTI Consulting	<p>For certain cases, project level de-duping can make it difficult to tie custodians to documents later in the process.</p>
9. GGO	<p>Most commonly, when deciding not to de-dupe across the project corpus, GGO clients cite an interest in preserving what GGO calls the social networking view - who is talking to whom about what, and when. Counsel sees value in having context (if needed) for the generation and use of duplicate documents. Furthermore, they believe that custodian-level de-duping adequately strikes the balance between reducing duplicative review effort and preserving useful data.</p>
10. Iris Data Services	<p>Many times this is done as a result of government requests to produce one version of each document for each custodian. While these can be re-populated when it comes time for production, it adds extra time and expense when it comes time to actually produce. Many case teams are so short on actual review time that they cannot afford this extra production step. Leaving the dupes in the review batches is also made easier by modern review tools, such as Iris' Unify product, that allows users to extend production decisions to all duplicates in the database. This allows case teams to de-duplicate within custodian but achieve review and production efficiencies similar to de-duplicating across the entire project.</p>
11. Kroll Ontrack	<p>The main reason why project level de-duplication isn’t a popular option is primarily centered on production sets. When it comes time to produce documents for custodians, project level de-duplication loses the context for where the document was found. Lawyers cannot fully state which custodians had which documents because they are processed sequentially. If custodian A had ‘document 1.doc’ and Custodian Z had ‘document 1.doc’, it would only be noted that the document was found under Custodian A.</p>

Company	Response to “Why don’t attorneys elect to use project-level deduping on all projects?”
12. LDM Global	The choice on how to de-dupe is really based on the case itself. For many cases, firms want to eliminate as much data as possible in order to cut down on review time, costs etc. There are however instances where reducing the data set as much as possible is not always an option. For example, some cases require production of only certain custodians (even if all custodians have been collected and processed). In these cases you want all custodians to have a full set (not excluding the duplicate files eliminated when compared to other custodians).
13. LDSI	Mostly this is to do with their review techniques. If they distribute allocate documents with a custodian level review, then documents will have been removed from specific custodians, which can create significant issues. Our recommendation to overcome this is to do issue type reviews rather than custodian ones.
14. Rational Retention	
15. Recommind	Concerned about missing something , usually resulting from lack of technological sophistication
16. StoredIQ	Scope of custodians may change during litigation. Reviewers need the context of each custodian.
17. Trilantic	US litigation requirements
18. Valora	They don't understand it. They don't want to think about the ramifications. They just want to go, go, go onto review. Sometimes they are concerned about documents "disappearing." It might be helpful for your article to talk about the difference between duplicate detection and actual removal. We use the terms "quarantine" and "remove/cull" to explain the difference.

Response to Question 12, Do you provide reduping services?

Company	Do You Provide Reduping Services	Have done so since:
1. Act	No	-
2. BIA	Yes	2002
3. CaseCentral	Yes	2008
4. Clearwell	Yes	2006
5. Daticon	Yes	2006*1
6. Encore	Yes	2003
7. Fios	Yes	2001
8. FTI Consulting	Yes	2001
9. GGO	Yes	2002
10. Iris Data Services	Yes	Company inception
11. Kroll Ontrack	Yes	1999
12. LDM Global	No	-
13. LDSI	Yes	2005
14. Rational Retention	-	-
15. Recommind	Yes	2007
16. StoredIQ	Yes	Via reproduction
17. Trilantic	Yes	2005
18. Valora	Yes	2007

Footnotes

*1 See Response to Item #13.

Response to Item #13, Any other thoughts or insights you'd care to offer on deduping as a cost-saving strategy or litigation risk avoidance strategy?

Company	Response to "Any other thoughts or insights you'd care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:
1. Act	Not at this writing
2. BIA	Cross-custodian de-duping for review purposes (not necessarily prod purposes) is critical to ensure consistency during review with respect to tagging by different reviewers and also to ensure consistency in redaction.
3. CaseCentral	<p>Culling duplicates is one example. But treatment of related documents in general -- be they duplicates, near-duplicates, attachments, threads or concepts -- can have enormous impact on cost and risk.</p> <p>For example:</p> <ul style="list-style-type: none"> - Bulk-tagging - CaseCentral's Red-line review (highlighting differences between near duplicates) - Comparing designations across related documents <p>Cost-saving and risk avoidance must include all types of related documents and common tasks, from culling to review and production. Tools that allow this, optimize eDiscovery and minimize cost and risk.</p>
4. Clearwell	The impetus for de-duping is most always a cost saving measure, however, we have found it's important to maximize deduplication during search, analysis, and review while at the same time managing litigation risk by always tracking all occurrences of each document within the case database. This also provides the greatest possible flexibility at production time if it becomes necessary to produce documents with different deduplication parameters (i.e., a horizontally-deduped production for inside counsel and a vertically-deduped one for opposing counsel).
5. Daticon	<p>* Daticon EED can provide "re-duplication" of Meta data about de-duplicated items as a report or as associated tags for loaded items. For example, a report will show all locations in which a particular file was stored in a file system or email environment. When applied as associated tags to loaded items, information about de-duplicated items is searchable and viewable in the Discovery Partner review workplace.</p> <p>We do not re-populate the physical files that were de-duplicated during initial processing. Generally speaking, the re-population of physical files at the time of production serves only to increase production cost and complexity while offering no obvious value in terms of produced content.</p>

Company	Response to “Any other thoughts or insights you’d care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:
6. Encore	Advanced approaches such as duplicate consolidation, thread consolidation and near de-duplication offer significantly improved results over basic de-duplication techniques. De-duplication helps ensure consistency of review decisions and reduces redundant review costs.
7. Fios	<p>Fios utilizes over 30 different standard approaches to de-duplication. Along with the court-order issue on project level de-duplication, Fios regularly applies other reduction techniques result in significant data reduction when project level de-duplication is not used, such as specific path isolation, custodian isolation for review, and evidence repositories for identified hot custodians.</p> <p>Fios also offers a wide range of professional services, independent of technology, that can help attorneys create more effective search term lists in advance of ESI processing and review. This advance planning has proven to cull down evidence sets by as much as 90%, therefore dramatically reducing the costs of e-discovery. Advance planning, content mapping and sampling of potentially relevant electronic evidence can also mitigate the risk of missing potentially relevant evidence and/or ensuring all ESI has been properly preserved and collected.</p>
8. FTI Consulting	The reduction percentages requested in 9A, 9B, 9C and 10 above are very difficult to provide generally, due to wide variance between project size, similarity of custodian datasets, filetypes, case subject matter, and many other possible factors that can occur on a case-by-case basis.

Company	Response to “Any other thoughts or insights you’d care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:
9. GGO	<p>As it relates to suppressing duplicative documents from review, generally, GGO's software methodology focuses the hosted review database to one unique document per custodian. All preserved documents and dupes can be returned to the review environment. GGO does not “remove” duplicates per se, rather GGO prevents duplicative review.</p> <p>GGO's proprietary review software automatically propagates reviewer marks/issue codes to all instances of the files across the project corpus – even if they are not viewable in the review environment. This 1) prevents collisions (different marks for duplicative documents); and 2) provides reviewers of different custodians with instant context for their own custodian review decisions. Preserved, un-viewable documents can be re-introduced with the mark intact, and produced even if the duplicative document was never viewed. In this way, we allows clients to automatically suppress duplicates (and Near Dupes - see below), while preserving the ability to track and audit every suppressed document on-demand, if need be.</p> <p>Presentation or suppression of duplicate documents in the review environment is a matter of attorney preference, and deployment of GGO's software is tailored to the review strategy for the matter. When gross relevancy review is carried out by contract attorneys, Suppressing duplicates across the entire corpus works well – less distraction and screen clutter. When attorneys are also issue coding documents, GGO (and its 150+ law firm and enterprise GC clients) have found that presenting only one dupe per custodian provides the most uncluttered and efficient review. As hot doc’s are encountered, attorneys can quickly sort on the duplication field and see all parties that are associated with the hot information.</p> <p>Another growing trend amongst GGO's software users is the use of its Near-Dupe and Thread Analysis functionality. Near-Dupes are defined as 1) message groups within an email thread, and 2) documents that are substantively exact, but not exact duplicates.</p> <p>GGO's software identifies Near Dupes based on email attributes, proprietary algorithms and client-customizable criteria. Near-Dupes are assigned a thread number and can be grouped, tracked and reviewed across the entire corpus and at the custodian level.</p>

Company	Response to “Any other thoughts or insights you’d care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:
10. Iris Data Services	<p>De-duplication mitigates risk as well as minimizes cost. Regardless of the implemented strategy (be it across the entire project or simply within custodians), de-duplication can be used to effectively minimize the risk of review mistakes, unintentional breach of privilege, and other costly errors. This is achieved by limiting reviewers to making one decision on identical documents as opposed to having several reviewers making multiple decisions on identical documents.</p> <p>Similarly, when used in early case assessment and/or prior to full processing, de-duplication can be an invaluable tool to reduce cost. De-duplication reduces the raw amount of data that requires eDiscovery processing and, more importantly, the number of documents that have to be reviewed.</p>
11. Kroll Ontrack	<p>De-duplication will not impact reduction as much as well defined keyword and date filter parameters. These in conjunction with custodian level de-duplication will give you the most reduction in size while allowing legal teams to produce by custodian.</p>
12. LDM Global	<p>The decision to de-dupe is primarily based on the case itself. Almost all cases take advantage of reducing the data population through de-duping but the method chosen is driven by case requirements. You should be aware of the production requirements or legal strategy prior to choosing a de-duping method.</p>
13. LDSI	<p>Deduplication is an essential process for nearly edd projects, but a number of factors need to be considered before deciding on the strategy. This includes considering the source of the data (ie, back up, live, email servers etc), size of custodian pool , review methodology and type of litigation and what the firm is being required to do (ie. witness statements, regulatory project, litigation where smoking guns are sought all require different dedupe strategies).</p> <p>Persons providing instructions need to understand the implications of their dedupe instructions and to understand hashing methodologies and that the choice of fields can have a large impact on duplicate inclusion, eg. internal fields such as internet message ID, when both notes and microsoft email is included, and of the time settings of emails from various international locations.</p>

Company	Response to “Any other thoughts or insights you’d care to offer on de-duping as a cost-saving strategy or litigation risk avoidance strategy:
14. Rational Retention	<p>Yes – we at Rational Retention believe that the entire concept of de-duplication is, or at least should be, an anachronism. By identifying, analyzing (including search) and holding in place it is possible to only collect a single copy of any file, regardless of how many copies are distributed across the enterprise.</p> <p>Along with the collected file should be a listing of the various locations where it resides. At review time the file and its objective, subjective and location metadata should be single instance stored and accessible by the reviewer as needed.</p> <p>Beyond the obvious costs saving for processing this approach greatly enhances automated remote collection. If files are identified for collection, but reside on a system not currently on the network; the e-discovery system should simply move on and collect from the next available location where the files reside.</p> <p>The above describes how Rational Retention works. We are always happy to discuss.</p>
15. Recommend	<p>We tend to work with larger, more sophisticated firms and clients, for whom deduping is assumed to be done with every case (and is considered older, more rudimentary technology). In fact, we work with an increasing number of clients who conduct extensive deduplication (what you refer to as Project-Level Dedupe) at the time of collection (using our software), well before outside counsel ever see anything.</p>
16. StoredIQ	<p>Email de-dupe needs to be configurable as requirements will vary from customer to customer. De-dupe with custodians is a cost effective measure especially when you can deliver duplicate counts to the reviewers.</p>
17. Trilantic	<p>No response.</p>
18. Valora	<p>Yes. De-duping is really only the tip of the iceberg. Any discussion of de-duping for the purpose of lowering processing, hosting or review costs should really include the full suite of population analysis tools: de-dupe, near-dupe, email thread group, issue-tag/categorization, presumptive priv and responsive/relevance tagging.</p>