

eDiscovery Institute



Report on
Kershaw-Howie Survey of E-Discovery Providers
Pertaining to
Email Threading

By

Anne Kershaw

anne.kershaw@akershaw.com

www.akershaw.com

and

Joe Howie

joe@howieconsulting.com

www.howieconsulting.com

Release Date: January 5, 2010

Contents

Overview of Results	3
Respondents	3
Email Threading Maturity	3
Metrics	3
Benefits of Email Threading	4
Specific Features	4
Pricing and Distribution	4
Other Technologies	5
Related Services	5
Respondents	6
Survey Responses	7
Product Identification	7
Metrics	8
How Threading Achieves Savings	9
Specific Features: Forwarded Emails, Text Verification, Different Systems, Tagging Threads, Paper-Copy Email	12
Specific Features: BCC's and Using Threading information in Other Review Platforms	13
Distribution and Pricing	15
Interaction of Threading with Other Technology	16
Thoughts on Cost-Effectiveness	19
Notes	21
Survey Invitations	22
Survey	23
Appendix – Terminology	26

Overview of Results

In September/October of 2009, Anne Kershaw and Joe Howie surveyed leading providers of electronic discovery software and services to determine what their experience had been with using email threading technology as a way of providing cost-effective e-discovery review. This report provides the results of the responses.

Respondents

The 13 respondents offered a mix of services and software. Most of them have developed their own email threading capabilities as part of their service offerings which includes processing and hosting. Three of them (Equivio, Clearwell and OrcaTec) are software providers with email threading offerings. Two of them are service providers which have offerings that include Equivio.

Email Threading Maturity

Email threading is a relatively new technology. Most of the respondents have been offering it for three years or less.

Metrics

One of the challenges of the survey was to try to isolate the impact of email threading from other efficiencies achieved through the use of complementary technologies or processes. For example many of the respondents also offer near-duping or concept clustering, both of which can speed review by grouping like threads together.

The best metric to gauge the potential impact of email threading by itself is the average number of emails per thread. In simplistic terms, the longer the threads the more time is saved by not having to examine earlier emails. Because the last email in a thread will be the longest one, one can't say that, e.g. if there are five emails per thread the reviewer will save 80% of the time that would otherwise be spent – the earliest emails are the shortest and less time is saved by being able to skip them. Assuming all of the emails added an equal amount of content to the thread, the last email in a thread of five emails would have 33 % of the overall content of the thread with a theoretical savings potential of 67%:

Hypothetical Avg. Emails per Thread =5 ¹	
Email position in thread	Amount of content (Earlier emails plus added content)
1	1
2	2
3	3
4	4
5	5
Total	15
Email #5 as % of Total Content (5/15)	33%
Theoretical savings potential in reviewing content	67%

¹ Similar calculations for average emails per thread of 2 and of 10 results in the following numbers:

	Avg. Email/Thread=2	Avg. Email/Thread=10
Content in last email as % all content	66%	18%
Theoretical potential savings in content review time	34%	82%

KERSHAW-HOWIE EMAIL THREADING SURVEY

The respondents reported that across all their projects, the average email per thread was 4.9 with average savings in review time of 36% which seems reasonable in light of the theoretical potential savings. Most of the respondents reported having had projects with average email thread size of 9 to 12 emails per thread with savings ranging up to 80%. They also reported projects with the lowest emails per thread of 1.5 to 2 emails per thread with the lowest savings in the range of 5 to 10%.

Benefits of Email Threading

The benefits of email threading can be summarized as follows:

- **Faster review** – earlier emails in a thread can be skipped because their content is repeated in a later email in the thread.
- **Consistency** – the reviewer has the option of treating all emails in the thread consistently.
- **Decision-making Quality** – reviewers can understand the context of the entire email conversation and therefore make more informed decisions on things like relevance and privilege.

Some of the specific features mentioned by various respondents:

- **Bulk tagging** – ability to tag all emails in a thread with one step.
- **Difference highlighting** – highlight the differences in content between two records.
- **Work assignment** – assistance in allocating complete threads to individual reviewers.
- **Visual or graphical representation of the thread.**
- **Flagging begin and end points in threads.**
- **Organization or sorting of emails in a thread.**

Specific Features

- **Forwarding.** Most but not all respondents indicated that their systems could associate a forwarded email with an earlier thread.
- **Content verification.** Most but not all respondents indicated that their systems would verify that the text or content of earlier emails was contained with the later emails. Equivio claims a 100% “no false positive” confidence that emails placed in threads from electronic ingest data actually belong in the thread.
- **Differing Email Systems.** Most but not all reported that their systems could associate emails from different email systems (e.g. Notes and Outlook) into threads.
- **Bulk Tagging.** Most respondents reported being able to tag all the emails in a thread with bulk tagging.
- **Paper-source Emails.** While most respondents reported that they could even associate emails that had been produced in paper form into threads, many of them noted that they could not do this as confidently with paper-source records as they could with electronic source emails – undoubtedly because of OCR errors and missing metadata values.
- **BCC’s.** BCC’s were treated as part of the thread. One of the respondents indicated it created a single unified BCC list.
- **Other Platforms.** Most of the respondents indicated that they could export a Thread ID-type field or fields to other review platforms.

Pricing and Distribution

There were a variety of pricing options offered by respondents, including “all you can process” software licenses, per GB, per email, per item, per custodian and per page fees options. One offered pricing based on “relevant” data. Equivio and OrcTec offered their technology in software development kits for integration in other platforms.

Other Technologies

Many of the respondents indicated that they offered other complementary technologies that improved review speed and/or review quality, including:

- Auto-Coding
- Automated First Pass Review
- Automated Privilege Detection
- Concept Clustering
- Deduping
- De-threading
- Difference highlighting
- Interesting Phrase Finder
- Key Phrase Identification
- Keyword Analytics
- Language Identification
- Near-duping
- Review Management

Related Services

Certain of the respondents offered consulting or services in addition to software.

Respondents

The following companies provided responses to the survey:

Company Name	Person Responding	Company Website
Anacomp, Inc.	Jeff Friedman, Sr. Product Manager	www.anacomp.com
Capital Legal Solutions	Mark Reichenbach, VP for Client & Industry Development	www.capitallegals.com
Clearwell Systems	VP for Client & Industry Development, Director, Product Management	www.clearwellsystems.com
Daticon EED	Scott Hipsman, Director, Service Design	www.daticon-eed.com
Equivio	Director, Service Design, Director, Service Design	www.equivio.com
InterLegis, Inc.	Director, Service Design, President	www.interlegis.com
Kroll Ontrack, Inc.	Michele Lange, Director of Legal Technologies	www.krollontrack.com
Logik	Adam Reilly, Director of Legal Technologies	www.logik.com
OrcaTec	Herb Roitblat, Principal	www.orcatec.com
Recommind	Craig Carpenter, VP of Marketing & General Counsel	www.recommind.com
TCDI	Tom MacKenzie, VP Business Development	www.tcdi.com
TRILANTIC	Nigel Murray, Managing Director	www.trilantic.co.uk
Valora	Sandra Serkes, CEO & President	www.valoratech.com

Survey Responses

Product Identification

Company Name	Name of Offering	Since	Source of Technology	
			Dev. Own	License From:
Anacomp	Anacomp eDiscovery Services / CaseLogistix	2008	No	Various Technology Partners including Equivio. Integrated with CaseLogistix.
Capital Legal Solutions	Part of our eZReview® product, termed "Email Analytics"	June 2008	Yes	
Clearwell Systems	Clearwell E-Discovery Platform	Jan 2006	Yes	
Daticon EED	CompareVue	January 2007	No	Our own development along with Equivio
Equivio	Equivio>EmailThreads	June 2007	Yes	
InterLegis	Discovery360 Reviewer	2000	Yes	
Kroll Ontrack	ETMetric ²	February 2009	Yes	
Logik	SimLogik, ThreadField	2008	Yes	
OrcaTec	OrcaTec email threading	2007	Yes	
Recommind	Insite Legal Hold; Axcelerate eDiscovery	2008; 2007	Yes	
TCDI	Internal Product - Not Named	January 2009	Yes	
TRILANTIC	Equivio	2007	No	Equivio
Valora	Email Thread Grouping & Email De-Threading (2)	2005	Yes	

Average Date First Offered Email Threading: 2006

² "This technology is available as a standalone component able to be embedded in a variety of electronic discovery applications to generate better, faster search results. We have also embedded it within our own document review platform, Ontrack Inview, for direct clients to utilize in the course of a discovery processing and hosting project."

KERSHAW-HOWIE EMAIL THREADING SURVEY

Metrics

The metrics section used the following definition of Project-Level Average Emails Per Thread (PLAEpT): The total number of Emails divided by the total number of threads (including single-email threads where there is no known subsequent Reply or Forward).

The directions also assumed that the savings comparisons compared reviews with threading to reviews where the collection had already been deduped across custodians.

Company Name	Project-Level Average Emails per Thread			Savings in review from Email Threading compared to no Email Threading		
	Q 4.1 Avg	Q. 4.2 Highest	Q. 4.3 Lowest	Q. 4.4 Avg	Q. 4.5 Highest	Q. 4.6 Lowest
Anacomp						
Capital Legal Solutions ³	4.2	11	2	10%	20%	5%
Clearwell Systems	5	10	2	30%	55%	10%
Daticon EED ⁴	4.8	11.4	1.7	58%	82%	
Equivio ⁵	3-5					
InterLegis ⁶	2.1	3.5	1.5			
Kroll Ontrack						
Logik	10-15					
OrcaTec ⁷						
Recommind	3	10	2	20%	50%	5%
TCDI	4.6	9.2	2.3	25%		
TRILANTIC ⁸	2.6					
Valora ⁹	6	12	1.5	75%		
TOTALS	48.8	67.1	13	218	287	30
NUMBER OF RESPONSES	10	7	7	6	5	4
AVERAGE	4.9	9.6	1.9	36%	57%	8%

³ On savings: "Clients Can't Officially Disclose - Anecdotally report..."

⁴ On average savings: "On average, CompareVue speeds review rates from 50/60 docs p/hr to 110/150 p/hr." To convert this to percentage savings we used calculations that compared 55 docs/hr without threading to 130 with threading.

On maximum savings: "The largest review rate we have found using CompareVue was >300 docs p/hr." We compared this to 55 docs/hour for before threading."

On minimum savings: "The lowest review rate we've seen using CompareVue is 32 docs p/hr."

⁵ "Based on information provided by our service provider partners."

⁶ Interlegis reported savings on average of up to 80% that resulted from using their combination of technologies with highest savings in excess of 80%.

⁷ "We provide software, we do not provide services."

⁸ "We have only analysed this on one project, the average was 2.6 and this was across a set which appeared to have very long chains."

⁹ On maximum savings: "100% (Fully automated. No human review.)"

On minimum savings: "0% (Client opted not to use/protecting billable hours)."

How Threading Achieves Savings

These are the responses to question 5:

“How does your email threading offering enable faster reviews?”

Q. 5, How does your email threading enable faster reviews?	
Anacomp	<p>Allow users to review unique data in context, efficiently and accurately.</p> <p>1) Efficiency of document review when reading documents within context. Assuming all documents within a thread would otherwise need to be reviewed, being able to group thread together and keep them as document family (not unlike parent / child relationships) will assist reviewers performing a detailed designation review. Documents read in context of their full thread can be done quicker (to the tune of several seconds to minutes / document) because the entire document story is presented rather than pieces of a story with missing chapters.</p> <p>2) Accuracy / consistency similar to above, if threads are maintained like families, then responsive and privilege designations can more accurately / consistently be maintained. If the core email to a thread is privileged, then a good argument could be made that the entire thread should follow the same designation (this is not to say threads will always share designations as a unit.) Likewise, as a thread is read in context of it's entire story, additional context can lead to more accurate use in the overall case. (reviewing each email individually would lose this benefit.)</p>
Capital Legal Solutions	<p>Entire conversational thread displayed, reviewer sees everything in-context, makes decision on the entire thread in-mass.</p>
Clearwell Systems	<p>Clearwell provides a "Discussion Thread view" of case data, allowing reviewers to review emails by Discussion Thread instead of only keyword hit by keyword hit. By linking together all related messages into chronological threads that capture the entire discussion, including all replies, carbon copies, blind carbon copies, and forwards, reviewers understand the context of evidence more quickly and make bulk decisions on entire threads vs. by a single document -- thus enabling a more rapid review of emails and attachments.</p>
Daticon EED	<p>Our CompareVue feature includes the use of a "Difference Viewer" that illustrates specific differences from the first reviewed thread member to others thread members. Reviewers are able to take a coding decision that was made against the first reviewed thread member and quickly test whether that same decision can be applied to each document in the email thread. By focusing the reviewer's attention on the differences between documents, the result is more document decisions per hour and vastly improved coding accuracy.</p>

Q. 5, How does your email threading enable faster reviews?	
Equivio	<p>The Equivio tool simplifies the review of emails, while allowing the review of the email within its original context:</p> <ol style="list-style-type: none"> 1. Assign email threads for review, ensuring coherent review of email data. 2. Start review of an email thread by reading the first "Inclusive". An Inclusive is the last email in a thread. Equivio verifies that each Inclusive contains the entire history of the thread (the determination of the "Inclusive" email is based solely upon the content of the email and not on the metadata. As such, the solution is robust allowing cross-platform processing). 3. Review the other Inclusives in the thread with the Equivio>Compare tool. Equivio>Compare highlights only the differences between emails, eliminating the wasteful review of email data that the user has already seen. The user reviews unique data only. 4. Users can bulk handle all the emails in a thread, reducing time and cost, while also ensuring that all the emails in the thread are treated consistently.
InterLegis	<p>Discovery360's interface allows reviewers to access an email thread through one click. Reviewers can also utilize advanced email analytics, such as:</p> <ul style="list-style-type: none"> - visual mapping of communication chains - visual mapping of communication timelines. - similarity matching of emails containing certain content, concepts and attributes - last email thread identification - email chain breaks and hidden content - subject matter assignment to matter-specific reviewers - and more. <p>Once relevant emails are found, reviewers can then make individual or bulk document decisions such as for relevancy, privilege, responsiveness, production, or to assign for additional review.</p>
Kroll Ontrack	<p>ETmetric™ e-mail threading tool allows document review teams to view all related e-mail, sent and received, in a conversation thread – reducing the number of e-mails that need to be reviewed.</p> <p>ETmetric™ identifies the "Start-Point" and "End-Point" of each e-mail chain and organizes the e-mail chronologically.</p>

Q. 5, How does your email threading enable faster reviews?	
Logik	ThreadField allows e-mails within a chain to be grouped together via a sortable field value. Related messages have common prefixes in a string which increases in length with each successive message. Therefore, ThreadField makes it possible to group and tag messages from a thread in almost any review system. Additionally, SimLogik assigns a group ID to near-duplicate document, which tends to find threads regardless of the underlying storage mechanism.
OrcaTec	Organizes conversations into one place. Let's people see information in context. If documents are assigned more or less randomly to reviewers, there is no opportunity to take advantage of the relationships among emails. Let's entire threads of emails to be tagged at once.
Recommind	By reducing the number of redundant messages which must be reviewed. Please also remember that our threading feature is part of our computer-generated review technology; as such, it is a relatively minor time-saving feature within our product offering.
TCDI	By grouping emails within a thread and identifying the most inclusive emails within the multiple branches of a thread, reviewers are provided all of the emails in a thread grouped together in one location for coding. One-click bulk coding can be performed up specific branches of a thread thus saving significant reviewer time.
TRILANTIC	The software identifies the email's which are complete or complete but without an attachment which existed earlier in a chain. Only these complete email's need to be reviewed.
Valora	By grouping email threads together, review personnel can see logical conversation chains, in any order that makes sense to them (chronological, reverse-chronological, by author, by topic and so on). Email thread grouping (ETG) is a lot like near-duping. Similar documents that have an associated context are viewed in a connected, logical manner. ETG can also help highlight missing information or intentional, but undisclosed, forwarding to parties. ETG also helps bring important attachments to the entire conversation, not just its origins.

Specific Features: Forwarded Emails, Text Verification, Different Systems, Tagging Threads, Paper-Copy Email

The following table has responses to questions about specific features/functionality:

- **Forwarded Emails.** Question 6: “Does your email threading associate Forwarded Emails with earlier emails in the thread?”
- **Verifying Text of Earlier Emails.** Question 8: “Does your email threading verify or indicate whether the text of earlier emails is included in subsequent emails?”
- **Emails from Different Systems.** Question 9: “Is your system able to identify threads when emails are exchanged between different email systems, e.g. Outlook, Notes, Groupwise, AOL, Gmail, etc. For example, if one custodian produces NSF emails and another produces a PST file, will you system combine emails from those two sources into common threads?”
- **Tagging Entire Thread.** Question 10: “If you provide a review platform, Can you tag all emails in a thread in one click or operation in that review platform?”
- **Threading Paper-source Emails.** Question 12: “Is your process able to include scanned copies of emails that had been printed out in its threading? For example, if a custodian had saved only printed-out copies of emails, could your system work with the OCR from the scanned images of those emails to place them in threads?”

Company Name	Q 6, Forwards Included in Thread	Q. 8 Verify Text Earlier Emails	Q. 9 Diff. Email Systems	Q. 10 Tag Whole Thread	Q. 12 Threading Paper-Copy Emails
Anacomp	Yes	Yes	Yes	Yes	Yes (not as confidently)
Capital Legal Solutions ¹⁰	Yes	Yes	No	Yes	No
Clearwell Systems	Yes	Yes	Yes	Yes	Yes (not as confidently)
Daticon EED	Yes	Yes	No	Yes	Yes (not as confidently)
Equivio	Yes	Yes	Yes	No Platform	Yes
InterLegis	Yes	Yes	Yes	Yes	Yes
Kroll Ontrack	Yes	Yes	Yes	Yes	Yes
Logik	No	No	Yes	No Platform	Yes (not as confidently)
OrcaTec	Yes	Yes	Yes	No Platform	Yes (not as confidently)
Recommind	Yes	No	No	Yes	Yes
TCDI	Yes	Yes	Yes	Yes	Yes (not as confidently)
TRILANTIC	Yes	Yes	Yes	No	Yes (not as confidently)
Valora	Yes	Yes	Yes	Yes	Yes

¹⁰ “Regarding item #12, from an automatic standpoint, this is a road map item.”

Specific Features: BCC's and Using Threading information in Other Review Platforms

Questions 7 and 11 dealt with BCC's and working with other platforms:

- **BCC's.** Question 7: "How does your email threading process BCC's? (e.g. Shows only the BCC copy if available or show the BCC copy as a related email or some other treatment.)"
- **Other Review Platforms.** Question 11: "What process do you have to permit working with threads in other review platforms such as Concordance, iCONNECT or Summation?"

Company Name	Q. 7, BCC Treatment	Q 11, Other Platforms
Anacomp	BCCs are processed as related to the overall thread 'family' and would show up appear to the reviewer in that context.	Anacomp provides CaseLogistix platform for review, but as necessary, can provide load files with thread data to all other major platforms.
Capital Legal Solutions	It shows the bcc version in the chain of the email thread	On export we provide a field users sort-on to view entire thread
Clearwell Systems	Yes. Clearwell automatically aggregates bcc's across duplicate messages that are identical except for their bcc recipient(s) into a single unified bcc list	Clearwell provides review capability directly. If another review platform is used, Clearwell provides a Discussion Thread ID that can be imported into other review platforms. This allows discussion thread to be reconstructed by the 3rd party review platform.
Daticon EED	Our implementation will show the BCC copy when it is available as part of the entire thread.	Conversion+ERA (Evidence Relationship Analysis) provides value-added capabilities to both standard and non-standard conversion workflows for electronically stored information (ESI) when delivering to LexisNexis Concordance or CT Summation. Conversion+ERA identifies groups of documents in three relationship type categories; exact duplicates, near duplicates, and email threads. In addition to the standard client-defined fields in the delivered load file, Conversion+ERA provides three additional fields identifying each relationship type. These additional fields enable the results of a search to be sorted by one or more of the relationships, resulting in the creation of document groupings that can improve the accuracy of document review
Equivio	The BCC will be shown as metadata.	We are fully integrated with almost all of the review tools including those mentioned in the question.
InterLegis	Discovery360 will show the BCC copy, if available.	InterLegis can input Conversation Topic or Conversation Index fields which allows us to provide load files for, or import data from other review platforms.

KERSHAW-HOWIE EMAIL THREADING SURVEY

Company Name	Q. 7, BCC Treatment	Q 11, Other Platforms
Kroll Ontrack	BCC copy e-mails are analyzed as any other e-mail and are processed into the same thread, if applicable.	ETmetric outputs all thread information through an easily consummable XML format. The XML format can then be used by any application to load files into any third party system including Concordance, Summation, iCONNECT etc.
Logik	BCC is only shown if it is available.	ThreadField and SimLogik both generate fielded values which are compatible with almost any review tool.
OrcaTec	BCC is only shown if it is available.	Export thread information
Recommind	BCC is only shown if it is available.	None of our customers would conduct review in such linear-only platforms, so the question is NA for us.
TCDI	We show bcc's as a part of the email thread in the same way cc's would be shown.	Our email thread processing identifies and generates multiple pieces of information about the individual emails and their relationship within the thread. This information can be outputted as fielded data that can be imported to other platforms and can be used as long as the other platforms have a way to leverage fielded information associated with email threads.
TRILANTIC	Not handled	The output can be used in any review platform
Valora	BCCs are brought into the Email Thread Group (ETG), like all other related threads, and then indicated as such.	We export the ETG information as a new field and field value for automatic load into such systems.

Distribution and Pricing

Questions 13 and 14 dealt with distribution and pricing:

- Distribution, Q. 13: “Please indicate how interested parties could obtain your services or technology:
 - Directly from us (i.e. the company responding to this survey)
 - Through resellers or partners.”
- Pricing Model, Q. 14: “Without giving specific prices, please indicate what pricing options are available to clients:
 - License Fee Plus Per GB processed (LF+PGB)
 - License Fee Plus Per Item Processed (LF+PI)
 - Unlimited processing with no per unit fees, restricted to certain CPU’s (UL CPU spec)
 - Unlimited processing with no per unit fees unrestricted as to CPU’s (UL)
 - Per Gigabyte fee only (PGB)
 - Per Item fee only
 - Other”

Company Name	Q. 13, Dist		Q. 14, Pricing						
	Direct	Partners	LF+P Gb	LF+ PI	UL CPU spec	UL	PGB	PI	Other
Anacomp		✓		✓		✓			We present a number of different models both directly as well as through service partners - flexible depending on the appropriate model for the client(s) / projects
Capital Legal Solutions							✓	✓	We frequently tailor our pricing models to accommodate our clients needs.
Clearwell Systems		✓	✓						
Daticon EED							✓		Typically our eDiscovery services such as processing, hosting and production are available on a per GB basis. However, we also offer services that are billed on a hourly or per page basis. Custom pricing options may also be offered depending on the project or client requirements.
Equivio		✓			✓		✓	✓	
InterLegis		✓					✓		Gigabyte fee charged on only relevant data and includes productions.
Kroll Ontrack.						✓			
Logik		✓					✓		
OrcaTec		✓							Unlimited use, per server pricing.
Recommind		✓				✓	✓		
TCDI						✓	✓		Depending upon the specifics of the project or the term of relationship with the client we would either charge by the GB or provide with unlimited processing and no per unit fees.
TRILANTIC							✓	✓	
Valora		✓							Per GB, per custodian, per page and per document (per email) are all pricing options.

Interaction of Threading with Other Technology

Question 15 asked how the respondent’s threading technology fit with other cost-saving or quality improving technology:

“How does your implementation of email threading integrate with other cost-saving technologies such as deduping, near duping, concept clustering, etc. (Please, no more than 400 words – longer replies will be truncated.)“

Company Name	Q. 15, Other technologies
<p>Anacomp</p>	<p>Threading is a tool - both from a technology as well as a service perspective. Like other tools the goal is to ensure the best tools are utilized to most efficiently provide client solutions. Tools like Near-Duplication, Advanced Culling / Analytics (including clustering and categorization) of large data sets, Detailed Reporting, True Native file review, simple user interfaces and granular security are among many technologies we can bring to bear with a client. Whether the client is faced with a cookie cutter document review requiring multi-tiered workflow or the need to provide defensible search / identification techniques of large (terabytes) data sets to prepare for discovery and then culling down into a manageable review set - Anacomp's services and technologies are flexible to meet specific challenges.</p> <p>Rather than try and present a one-size-fits-all solution, Anacomp's implementation of these tools allow clients to develop processes that make sense for them.</p>
<p>Capital Legal Solutions</p>	<p>Seamless Integration.</p> <p>De-dupe, near-duping, clustering, and keyword analytics all are seamlessly integrated into our eZReview® platform</p>
<p>Clearwell Systems</p>	<p>Threading is fully integrated with deduping, so that only a single occurrence of each message in the thread is presented to the user (Clearwell tracks all occurrences of the email in the background for reporting, auditing, and downstream production purposes). Near-duplication is not integrated into email threading, since email threading is effectively a sort of near-deduplication for emails and it doesn't provide much additional value to combine the two approaches, in our opinion.</p>
<p>Daticon EED</p>	<p>CompareVue, a key feature of the Discovery Partner hosted review system, uses advanced analytics to group documents in a way that dramatically reduces review costs while increasing accuracy. Daticon EED’s CompareVue systematically organizes documents into clusters based on relationship and similarity of content (rather than on topic). Unlike review platforms that rely on topical or concept clusters as the principle mechanism for organizing documents, the CompareVue clustering model enables review acceleration when document-by-document review is required.</p> <p>Large data collections generally have a high occurrence of documents and email messages with very similar content. CompareVue leverages this fact by grouping near duplicate documents and members of email conversations. The real benefit provided by CompareVue, though, goes beyond just simple grouping. Documents in near duplicate clusters and email conversations, by definition, share a high percentage of content. With CompareVue, reviewers focus on complete review of a single item in the group of documents represented by a near duplicate cluster and an email conversation followed by a review of differences between the single item and members of the near duplicate cluster and email conversation.</p> <ul style="list-style-type: none"> • Once a coding decision is made for any member of a CompareVue group, the only factor that could change the coding decision for another member of the group is a difference in document content. CompareVue enables reviewers to quickly and visually navigate

KERSHAW-HOWIE EMAIL THREADING SURVEY

Company Name	Q. 15, Other technologies
	<p>through the document differences to confirm coding decisions.</p> <ul style="list-style-type: none"> • Focusing on the differences between documents, rather than complete review of all the content of each document, eliminates review of redundant content – saving time and money. Using CompareVue, average review rates will generally exceed 100 documents per hour per reviewer with the highest review rates exceeding 300 documents per hour per reviewer. • Reviewing all the documents related through near duplication and email conversation in a single session increases the accuracy and consistency of the review – again reducing review cost and time. In traditional “linear” review processes, similar documents may be reviewed at different times by different reviewers and, as a result, may receive different coding treatment. <p>In addition to facilitating effective review of near duplicate groups and email conversations, CompareVue automatically manages exact duplicate groups when clients choose no de-duplication or de-duplication within custodian. Through the CompareVue reviewer interface, any member of a duplicate group is accessible with a single mouse click.</p>
Equivio	<p>Equivio>EmailThreads is a stand alone product. It works well in conjunction with additional time and cost-saving techniques including deduping and near duping which Equivio offers as part of its solutions as well as additional technologies on the market.</p> <p>Equivio is built for straightforward plug-in as an OEM component, using the Equivio SDK, or as a standalone application, using the Equivio GUI, integrated within a broader third-party solution or service offering.</p>
InterLegis	<p>Discovery360 Reviewer is an all-in-one review platform, including the most effective and efficient technologies available, including email threading, similarity matching, near-duplicate detection, concept analysis, complex relationship analysis, as well as, review management tools to provide clients with a comprehensive and efficient review of their document collection. InterLegis provides the technology to reduce review time and a simplified pricing model to minimize and manage costs.</p>
Kroll Ontrack	<p>Ontrack Engenium offers a complete suite of products targeted specifically towards electronic discovery</p> <p>Ontrack Engenium Search Technologies</p> <p>Advanced search technologies such as concept searching, topic grouping, near duplicate identification and e-mail threading are demanded by cost-conscious discovery clients. Armed with Ontrack Engenium engines, your company will exceed clients' efficiency expectations, develop long-term client satisfaction and gain competitive advantages in this dynamic marketplace.</p> <p>Concept Searching</p> <p>Concept searching retrieves documents that are conceptually related to the search term rather than merely exact keyword matches. With concept searching, Ontrack Engenium allows users to rapidly find information that normally would be time consuming or impossible to find with keyword searching alone.</p> <p>Topic Grouping</p> <p>Topic grouping allows attorneys to learn valuable information about their case before reviewing the first document, providing an invaluable early case assessment tool as well as facilitating an accurate and consistent review. Ontrack Engenium speeds the review and navigation of large document sets by automatically grouping and labeling documents by subject matter - without the need for a query.</p> <p>Near Duplicate Identification</p> <p>Near duplicate identification allows document reviewers to detect documents that differ</p>

KERSHAW-HOWIE EMAIL THREADING SURVEY

Company Name	Q. 15, Other technologies
	<p>from each other by only a few words or paragraphs, irrespective of the file type, size or format. Ontrack Engenium's technology can specify the near-duplication similarity threshold. The engine also identifies a "core" document that is the most representative of the near-duplicate document set, allowing a reviewer to quickly determine whether review of the entire set is necessary.</p> <p>E-Mail Threading</p> <p>E-mail threading allows document review teams to view all related e-mail, sent and received, in a conversation thread - reducing the number of e-mails that need to be reviewed. The software identifies the "start-point and "end-point" of each e-mail chain and organizes the e-mail chronologically.</p>
Logik	<p>It provides methods for making data less amorphous by providing some basic structure. If documents can be reviewed with some logical order, rather than as a random stream, reviewers can avoid wasted time associated with swapping context between unrelated documents.</p>
OrcaTec	<p>We also offer near duping, concept clustering, concept searching, language identification, and interesting phrase finding.</p>
Recommind	<p>It is integrated into our advanced, non-linear review platform which includes automated grouping (concepts, language, MIME type, file type, custodian, key phrase, etc.) as well as our automated First Pass Review process and One Click Coding feature. Email threading is quite useful for those customers who only use our linear capabilities (a minority), but is less useful for those who use our automated capabilities (the vast majority) as we utilize the far more powerful tools referenced above.</p>
TCDI	<p>Our email thread capabilities were built to work independently or in conjunction with other technology to pull information together or push information out. For instance, if a user is conducting analysis based on concept clusters and finds specific emails of interest, one click can expand the record results to include all members of an email thread. Additionally, if all duplicates of an email attachment are deemed to require the same coding even outside of the email thread context, that coding can be automatically pushed out to all duplicates with the database. We are currently developing other integrated aspects of time-saving technology for leveraging during case assessment and review stages.</p>
TRILANTIC	<p>It is a separate process which can be used alongside other processes such as near duping. The results are stored in a set of fields in the database.</p>
Valora	<p>Valora's ETG is completely integrated with exact dupe detection (whether culled out, quarantined or just tagged), near-duping, and concept clustering, as well as AutoCoding (objective, subjective) and AutoReview capabilities: privilege detection, responsive detection and custom issue tagging. Email thread grouping is part of our standard offering.</p> <p>ETG also integrates with automated document coding so that customers can skip the processing, hosting and review of "internal" parts of the email chain, if so desired. For heavily redundant or similar email populations, this can be a major cost savings.</p>

Thoughts on Cost-Effectiveness

Question 16 provided an the respondents an opportunity to provide their thoughts on cost-effectiveness in the legal review process:

Any other thoughts you’d like to share on cost-effectiveness in the legal review process? (Please, no more than 200 words – longer replies will be truncated.)

Company Name	Q. 16, Cost Effectiveness in the Legal Review Process
Anacomp	Control through measurement. Legal review, is an exercise in controlling a process. Working with clients to set processes for every stage and reports to monitor and measure progress continually as well as at decision points. Cost Effectiveness is about controlling the data, controlling the process, and allowing clients to make smarter, more predictive, decisions based on a measured view of the data.
Capital Legal Solutions	As an industry thought-leader, Capital Legal Solutions understands that successful eDiscovery projects require three critical components: People, Process and Technology. CLS can ‘dial-in’ a solution comprised of any or all of those to deliver the best, most cost-effective result for the needs of your most demanding e-Discovery matters. We have deep litigation support experience with many members of our organization having held management positions with AmLaw 100 law firm’s litigation departments totaling decades of “in the line of fire” experience. CLS technologies like eZReview® are world-class with an unparalleled development and support team that delivers reliable, effective software built specifically for the demands of critical matters like yours.
Clearwell Systems	Since emails are the large majority of ESI types collected (typically 60%-75%) and eventually reviewed, Discussion thread technology is a critical component to early case assessment, culling and the ultimate review process. We’ve seen an increase of as much as several hundred more document decisions per hour when reviewing emails by Discussion Threads vs. topic clustering.
Daticon EED	In most matters, the largest area of cost is in document review. In order to reduce cost in this area, some vendors offer technology and some consulting services. At Daticon EED, we believe the best approach is a combination of the two and as such offer our clients an Expert Evidence Analytics Service Daticon EED’s Expert Evidence Analytics Service is focused on primarily electronic document collections, and utilizes a unique combination of software tools and analytics services in a fully documented offering designed to defensibly decrease the time spent in review. Bringing together our expertise, experience and a continuously expanding suite of technologies, a team of data analysts, search consultants and review managers collaborates with clients to fully understand the case, review team structure, and capacity. The Expert Evidence Analysis Service combines strategies developed by our experts with cutting edge technology and on-going consultation with the client and lawyers, utilizing: Expert case analytics, Expert culling services, Review optimization, Review team planning and training, Advanced review technologies. In addition, Daticon EED also maintains partnerships with both domestic and off shore review service providers in order to provide cost effective staffing solutions to clients.

KERSHAW-HOWIE EMAIL THREADING SURVEY

Company Name	Q. 16, Cost Effectiveness in the Legal Review Process
Equivio	By capturing and reconstructing email threads, Equivio eliminates data redundancy, exposing the unique data. The result is not only significant cost savings, but also reduced risk. Equivio technology directs the reviewer to what's important; viz. to documents that contain unique information, and to the unique information within each document. In addition, the grouping of the email thread enables the consistent treatment of documents. These are critical factors in reducing risk exposure for both the corporation and the law firm.
InterLegis	Cost-effectiveness in legal review begins at the onset of a project. Individual technologies, such as email threading, will help reduce costs and provide some efficiency. However, review teams need to consider all the phases of the discovery life cycle and utilize partners that can provide the proper solutions and consistency throughout the project. To that end, Discovery360 provides all available technologies in the industry today within a single, easy-to-use platform which can carry review teams from collection, through culling, processing, workflow, review, and production.
Kroll Ontrack	
Logik	
OrcaTec	
Recommind	The future of legal review is not in linear review, but in non linear review.
TCDI	
TRILANTIC	Email threading is one of those tools which captures the lawyers' imagination as they face daily challenges when reviewing eMails. I predict that, in time, this will become de facto in all Email data projects.
Valora	<p>Email thread grouping (ETG) is a natural complement to near-duping. The goal of both procedures is to bring like-minded documents together, whether the similarity is in content (near-duping) or in communication (ETG). In fact, many ETG groups will also contain near-dupes, so it is important to be able to easily identify the distinction.</p> <p>The best way to integrate ETG into a review process, though, is to utilize ETG (and the other techniques above) to fully pre-process the documents for dupes, privilege, responsiveness and issues, at the same time you are applying the ETG.</p>

Notes

Company Name	Q. 17 Notes
Anacomp	
Capital Legal Solutions	Regarding item #12, from an automatic standpoint, this is a road map item. However CLS processing and engineering offerings can leverage OCR and fielded data to obtain efficiencies.
Clearwell Systems	It is important to note that Discussion Thread technology differs widely amongst e-discovery vendors. For example, in contrast to other solutions that create threads solely on meta-data Clearwell builds threads on meta-data + content. This technology builds more complete threads (such as including emails when email content might be copied and pasted into another email). Clearwell's technology also detects news emails that were added from another email system (often when emails are sent externally from the set of collected ESI and sent around another company's email system).
Daticon EED	
Equivio	<p>Regarding question 4 -- Equivio is a software vendor. We do not perform the projects directly for service providers or end users. The data that we have been able to provide in answer to these questions is based on aggregate data provided by our partners and customers. As such, we do not have information relating to specific outlier statistics (highest, lowest) on individual projects.</p> <p>Regarding question 12 on OCR ingest data -- the email threads reconstructed by the Equivio technology have a 100% confidence level (no false positives), as for electronic ingest data. However, due to OCR errors in the input data, the "parent-child" relationships are sometimes not complete and the email threads may be truncated. However, where the thread cannot be grouped in a "parent-child" structure, our email thread technology works in tandem with our near-duplicate technology to group related emails in "sibling" structures.</p>
InterLegis	
Kroll Ontrack	
Logik	Review time savings were difficult to estimate as we do not have direct contact with reviews.
OrcaTec	
Recommind	
TCDI	
TRILANTIC	
Valora	<p>Valora has 2 email threading options that might be of interest. First is Email Thread Grouping (ETG), which is the conversation reconciliation type described in this survey. Second is Email De-threading, in which strung-together conversation threads are actually separated back out into their component inputs. This allows individual threadlets to be analysed and assessed individually for privilege, responsiveness, etc. Clients sometimes find this practice useful for preparing Privilege Logs.</p> <p>Valora wishes to stress that all capabilities discussed here, including ETG, are available for both paper & electronic documents. Valora supports printed paper email documents for full threading or de-threading. ETG, near-dupe, privilege, responsiveness, bib coding and issue coding are all available on paper docs, electronic files or mixed-media populations. Similarly, threading/de-threading capabilities are also available for text/instant messages (IMs), Twitter tweets and Bloomberg system messages.</p>

Survey Invitations

Contents of Email sent to survey respondents:

Subject: Invitation to Participate in Survey for Law Technology News Article

SENT ON BEHALF OF ANNE KERSHAW AND JOE HOWIE

We are inviting you to participate in a survey on the use of email threading technology as a way to provide cost-effective e-discovery review. The survey results will be used in an upcoming *Law Technology News* article on threading as implemented by leading e-discovery software and services providers. A more detailed compilation of survey results will also be posted on the website of the EDiscovery Institute.

The purpose of the article is to provide metrics around the type of savings that could be expected using this concept, recognizing that every collection is to a certain extent unique and may produce results outside the expected range. We believe that the net effect of the article will be to increase the proportion of cases or projects that are processed using this technology.

The survey is attached in PDF form. To participate, merely complete the form and click on the submit button – a copy of the completed questionnaire will be emailed to Joe Howie. We would appreciate it if you could submit the form by September 29, 2009.

Please let us know if you have any questions or comments or if you would like us to direct this invitation to some other person at your company.

Thank you in advance for your participation.

Regards,

Anne Kershaw and Joe Howie

The E-Discovery Institute

www.eDiscoveryInstitute.org

About Anne Kershaw

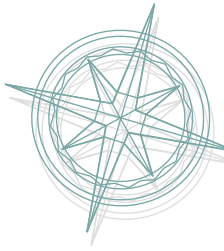
Anne is an attorney and senior consultant at A. Kershaw, PC//Attorneys & Consultants and the a co-founder of the EDiscoveryInstitute.org, a non-profit organization which researches the effectiveness of various litigation review technologies and processes. She teaches at Columbia and Georgetown Universities, 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, ACC, AAJ, DRI and others.

Survey

This is the actual survey:



eDiscovery Institute

Kershaw-Howie Survey on Cost-Effective E-Discovery Review: **Email Threading**

We’re preparing an article on the use of email threading technology as a way to help achieve cost-effective e-discovery review. The article will be submitted for a forthcoming issue of *Law Technology News* with a more detailed report being available on the website of the eDiscovery Institute.

The purpose of this survey is to gather information from leading e-discovery software and services providers about their actual experience using email threading technology. The results will help establish common industry benchmarks around the use of email threading as a way to speed review and minimize litigation risk.

To participate, merely fill in this PDF form and click on the “Submit” button – this will cause a copy of the completed form to be emailed to Joe Howie.

Please be sure to indicate if you’d be willing to be interviewed for the article. Also, note that we have included definitions in the Appendix to this survey.

Thanks for your participation!

Anne Kershaw

Joe Howie

Respondent Information	
Company Name:	<input type="text"/>
Company Website:	<input type="text"/>
Name of Individual Providing Response:	<input type="text"/>
Title of Individual Providing Response:	<input type="text"/>
Email Address of individual Providing Response:	<input type="text"/>
Phone Number:	<input type="text"/>
Are you willing to be interviewed for this survey?	<input type="radio"/> Yes <input type="radio"/> No

Your Company's Email Threading Offering	
Use the Notes area at the end of the survey if you would like to expand or explain any particular answer.	
1. What is the name of your Email Threading product or service offering?	<input type="text"/>
2. When was it first offered?	<input type="text"/>
3. Did you develop the technology or do you license the technology from another vendor?	<input type="radio"/> Developed Our Own <input type="radio"/> Licensed From <input type="text"/>
Metrics	
4. For the next six questions, assume that deduping across custodians has taken place <i>before</i> email threading.	
4.1 What is the average "Project-Level Average Emails per Thread" across all projects you've done? (See definitions, page 6.)	<input type="text"/>
4.2 What is the highest "Project-Level Average Emails Per Thread" you've seen in any individual project?	<input type="text"/>
4.3 What is the lowest "Project-Level Average Emails Per Thread" you've seen in any individual project?	<input type="text"/>
4.4 On average what percentage of review time for emails and attachments is saved by virtue of having threading available?	<input type="text"/>
4.5 What is the highest percentage of review time for emails and attachments that has been saved in an individual project as a result of having threading available?	<input type="text"/>
4.6 What is the least percentage of review time for emails and attachments that has been saved in individual projects as a result of having threading available?	<input type="text"/>

[Survey continued next page.]

Your Company's Email Threading Offering	
Overview of How Benefits Are Achieved	
5. How does your email threading offering enable faster reviews?	
Unique Features	
6. Does your email threading associate Forwarded Emails with earlier emails in the thread?	<input type="radio"/> Yes <input type="radio"/> No
7. How does your email threading process BCC's? (e.g. Shows only the BCC copy if available or show the BCC copy as a related email or some other treatment.)	
8. Does your email threading verify or indicate whether the text of earlier emails is included in subsequent emails?	<input type="radio"/> Yes <input type="radio"/> No
9. Is your system able to identify threads when emails are exchanged between different email systems, e.g. Outlook, Notes, Groupwise, AOL, Gmail, etc. For example, if one custodian produces NSF emails and another produces a PST file, will you system combine emails from those two sources into common threads?	<input type="radio"/> Yes <input type="radio"/> No
10. If you provide a review platform, Can you tag all emails in a thread in one click or operation in that review platform?	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> We don't offer a review platform
11. What process do you have to permit working with threads in other review platforms such as Concordance, iCONNECT or Summation?	
12. Is your process able to include scanned copies of emails that had been printed out in its threading? For example, if a custodian had saved only printed-out copies of emails, could your system work with the OCR from the scanned images of those emails to place them in threads?	<input type="radio"/> Yes <input type="radio"/> Yes, but not with same degree of confidence as the threading of electronic source emails. <input type="radio"/> No

Appendix – Terminology

- **Emails:** messages sent via the internet from one sender to one or more recipients, with provisions for a subject line, a message body, date/time sent, and From, To, CC and BCC fields; for purposes of this survey, the term “Emails” does not include IM’s or Tweets.
- **Email Address:** the string of characters used to send Emails to intended recipients in the form [UserName@DomainName.TopLevelDomain](#).
- **Email Address Name Label.** The string of characters used to describe the sender or a recipient of an email in addition to the Email Address, e.g. “Joe User <UserName@DomainName.TopLevelDomain>”; in some email systems the Email Address Name Label may be shown without showing the Email Address; different Email Address Name Labels may be correlated to the same Email Address by different email users, e.g. Bob Hart may have Joe User in his contacts list as “Joseph User” while Sue Foley has Bob Hart in her contact list as “Joe User”.
- **BCC Recipients:** Recipients whose Email Addresses are listed in the Blind Copy field; the Email Addresses of such recipients may not be shown on certain copies of the Emails.
- **Initiating Emails:** Emails that are sent to recipients without being a Reply to or a Forward of an earlier email.
- **Reply Emails:** emails that are sent by a recipient of an earlier email as a result of that recipient using the “Reply” or “Reply All” button or feature on certain email software or interfaces; normally the recipient of a Reply email is the sender of the earlier email and the recipients of Reply All emails are the sender and the disclosed recipients of the immediately preceding email; Reply Emails may or may not contain the text of the email to which the reply is being made.
- **Forwarded Emails:** emails that are sent by a recipient of an earlier email as a result of using the “Forward” button or feature on certain email software or interfaces; the recipients will usually include those whose email address did not appear on the earlier email or emails.
- **Email Thread:** all of the individual emails that were Replies or Forwards of an Initiating Email or subsequent Replies or Forwards; sometimes called a “conversation”.
- **Embedded Emails:** emails the text of which is shown in subsequent Replies or Forwards along with certain descriptive information such as From, To, CC, BCC, Date/Time and Subject.
- **Orphan Emails:** emails which appear to be Replies or Forwards but for which the initiating email cannot be identified.
- **Missing Emails.** Emails that do not exist as separate objects but which are quoted in one or more subsequent emails in a thread.
- **Branch End Email:** An Email to which there is no Reply or Forward; an Initiating Email to which there is no Reply or Forward is in a single length thread where the Initiating Email is also the Branch End Email; note that an Initiating Email may spawn multiple branches in the thread.
- **Project-Level Average Emails Per Thread:** The total number of Emails divided by the total number of threads (including single-email threads where there is no known subsequent Reply or Forward)