

The eDiscoveryJournal Report:

Early Case Assessment (ECA) Comes Into Focus in 2011

January, 2011

By:

Barry Murphy & Greg Buckles

CONTENTS

Executive Summary	3
Rising e-Discovery Costs Created A MisUnderstood Term	4
What Is Early case Assessment?.....	4
First-Generation And Next-Generation ECA.....	9
What Does “Analytics” Really Mean?	12
ECA - Diverse Users and A Variety of Use-Cases	17
ECA Anecdotes From The Trenches	20
Taking Corporate Control Of Litigation.....	20
Using Intelligence To Bring Efficiency To E-discovery Processes.....	21
Moving Forward With ECA.....	24
Appendix A - Processing, Analysis, and Review Features.....	26
Appendix B - Identification, Collection, Preservation, and Production Features	41

Disclaimer:

eDiscoveryJournal is not a law firm. All expressed opinions and content are provided for general educational purposes only and are not specific legal advice, even if the author is a practicing attorney. Neither eDiscoveryJournal nor the information contained herein should be used as a substitute for competent legal advice from a licensed professional attorney in your state.

eDiscoveryJournal believes reasonable efforts have been made to ensure the accuracy of all eDiscoveryJournal original content. Content may include inaccuracies or typographical errors and may be changed or updated without notice. All eDiscoveryJournal original content is provided “AS IS” and while we endeavor to keep the information up to date and correct, we make no representations or warranties of any kind, express or implied, about the fitness for a particular purpose, completeness, accuracy, reliability, suitability, or availability with respect to the information, products, services, or related graphics for any specific purpose. Any reliance you place on such information is therefore strictly at your own risk.

In no event will eDiscoveryJournal or any of its contributors be liable for any direct, indirect, punitive, incidental, special, or consequential damages or damages for loss of profits, revenue, data, down time, or use, arising out of or in any way connected with the use of the document or performance of any services, whether based on contract, tort, negligence, strict liability or otherwise.

EXECUTIVE SUMMARY

The term Early Case Assessment (ECA) means different things to different people. As a result, there is confusion and consternation around the term that stems from the fact that anyone can and everyone will claim to offer ECA functionality - archiving vendors, review vendors, search and collection vendors, legal hold vendors, and even ECA specialist vendors. The very name contributes to the confusion and although Early Data Assessment is a more accurate phrase, we will use ECA because that is how it has been branded in the market. The reality is that all providers can offer some level of ECA, which means that buyers need to know exactly what level of ECA they are looking for and what processes ECA will support. In the near-term, the onus is on the buyer to wade through the market and determine exactly which solution matches their requirement. This document will help buyers understand the functionality across the spectrum of ECA, use-cases of ECA, and create frameworks with which to evaluate ECA solutions.

RISING E-DISCOVERY COSTS CREATED A MISUNDERSTOOD TERM

Rising volumes of information and more stringent rules about the preservation of electronically stored information (ESI) drove eDiscovery costs to problematic levels in the early-to-mid 2000's. At the time, the highest eDiscovery costs organizations faced were in legal review fees and outsourced data processing (EDD). Because organizations were under the gun to collect and produce ESI within acceptable timeframes, over-collection was typical; as a result, too much data was sent out for EDD processing and legal review. The norm has been to simply hand over all collected data to a service bureau for processing. But, to process and review just 1 terabyte (TB) of data can cost in excess of \$1.5 million. Putting that into context, 1 TB of data can be the email and desktop content of approximately 10 typical corporate custodians. In addition to those processing costs, document review by law firms and other providers could be as high as \$500 per hour (or more, depending on the law firm). The rising costs gave solutions providers an opportunity to create tools and services that allow organizations to avoid or reduce eDiscovery costs.

And so was born a category that now drives many in the industry mad - early case assessment (ECA). The confusion around ECA stems from the fact that anyone can and everyone will claim to offer ECA functionality - archiving vendors, review vendors, search and collection vendors, legal hold vendors, and even ECA specialist vendors. The very name contributes to the confusion and although Early Data Assessment is a more accurate phrase, we will use ECA because that is how it has been branded in the market. The reality is that all providers can offer some level of ECA, which means that buyers need to know exactly what level of ECA they are looking for and what processes ECA will support. In the near-term, the onus is on the buyer to wade through the market and determine exactly which solution matches their requirement. This document will help buyers understand the functionality across the spectrum of ECA, use-cases of ECA, and create frameworks with which to evaluate ECA solutions. Despite the backlash against the term, ECA can have a profoundly positive business impact.

WHAT IS EARLY CASE ASSESSMENT?

The DuPont Legal Model was one of the first to truly bring up the aspect of early assessments of matters. At “the heart of the Legal Model process is an early assessment of the merits and economics of a case. If it is not worth pursuing or not winnable then resources will not be wasted on it.”¹ What the DuPont Legal Model did was shine a light on the cost savings that ECA could

¹ “*Inside the DuPont Legal Model*,” New Legal Review, May 11, 2010

potentially deliver. As the market evolved, the more traditional definitions of ECA focused purely on inventory and reporting on potential sources of unstructured ESI to derive an estimate of the potential cost of discovery. These definitions cover part of the “Tactical Scope Assessment,” but completely miss the initial “Strategic Assessment” that defines the fact pattern and potential key evidence that counsel needs to make a liability judgment. Senior management usually wants to know what likely really happened and what proof there is before they start spending money on software or services to scope the potential discovery costs.

Strategic Assessment Functions
Goal: support fact-finding investigations and liability assessments
Minimum features:
Index and Boolean Search – in place or at least without full explosion
Preview or Review of Native ESI – ability to view items and select or mark key items
Inventory or Catalog Reports – scan and report on potential network shares, desktops or collections so that investigators can focus on relevant areas
Custodian or User Reports – extraction of user names, email addresses, etc. to support investigation interviews and searches

There are many more features, especially content analytics, which add value during the strategic assessment; those listed above are the minimum to support the goal.

Tactical Assessment Functions
Goal: to define and quantify the scope and characteristics of potential eDiscovery to support budget decisions, Rule 26(f) scope negotiations, and other decisions in the identification phase of eDiscovery; to find the amount of potentially responsive information to understand the cost implications and then make decisions based on that

Tactical Assessment Functions
Minimum features:
Summary and detail reports on potential ESI sources.
De-duplication of ESI content
Extraction of domain, custodian, source and chronology characteristics to enable user to define and document relevant potential scope
Search term and criteria break down analysis, including term frequency, term clusters, directionality (inbound vs outbound) and other search relevance criteria for negotiation support. Term or hit frequency is critical to support arguments about overly broad criteria
Sampling searches to document exclusion of non-relevant sources
Cost projection tools that enable user to use known or estimated cost factors to calculate projected resources required, time and cost of potential collection criteria through discovery lifecycle

In considering both the strategic and tactical assessment goals, it is clear there is not a generally accepted, common definition of ECA – that’s part of the challenge that organizations have in determining what kind of ECA initiatives are right for them. An alternative is to think of ECA is as part of a spectrum of legal decision support. The goal is quite simply to use information to make fast decisions about matters in an attempt to save money. Confusion arises when vendors offer various levels of legal decision support and all of these vendors call their products ECA. While the vendor claims are theoretically true – many features can offer the ability to make early assessments about matters – the level of true assessment capabilities varies widely.

At the low end of the spectrum is simple search and filtering. The term simple may have negative connotations, but in fact, the ability to search and filter through information can provide huge savings in eDiscovery. Too often, organizations with tight deadlines turn over all collected

data for EDD processing. It's possible to drastically reduce processing costs by culling the data (searching and filtering and deduplicating) at the point of collection. Let's assume a small case with only 20 custodians, each of whom contribute 10 GBs of data. If an organization turns that data over for processing at the average cost of \$1,000 per GB, the processing alone will cost \$200,000. And remember, that's just the cost of processing. Having a law firm review that data just for responsiveness and privilege will cost more:

- There are typically 5,000- 10,000 (assume 7,000 for this example) documents per GB
- Average linear review rate is 60 documents per hour
- Average blended hourly review rate is \$150 per hour

Therefore, the cost of privilege review for 200 GBs of data will be about \$3.5 million. Processing and review will total close to \$4 million. That's a significant cost.

ECA solution providers claim the ability to reduce data collections by up to 80% – 90%. Factoring that down to just 50% to be conservative, that is still a potential savings of almost \$2 million. The reason this type of simple functionality can be called ECA is because it allows organizations to make case assessments based on the economics of eDiscovery. If eDiscovery for a case is going to cost \$2 million, but the case can be settled for \$25,000, the organization can quickly make the decision to settle the case. These types of savings are why ECA is something organizations can't ignore.

At the high-end of the legal decision support spectrum is functionality that helps reviewers actually determine case strategy and merits faster by grouping content into conceptual clusters, or by automatically determining the most potentially important documents, or by predictive coding of documents to make review faster. The key to most of these more advanced analytic systems is the ability to extract characteristics across the collection to effectively enable the documents to tell reviewers what the documents are. The more traditional tools expect you to know what you are asking for. This type of functionality helps organizations make faster decisions based on the merits of case rather than the simple economics (estimates of processing and review costs).

The question every organization should ask itself is, "what level of legal decision support do I need?" While every organization is different, there are some general guidelines:

All organizations should have, at a minimum, the ability to search, filter, and deduplicate all collected information, whether on-premise or through some kind of non-volume based service. Multiple applications will offer this natively (for example, email archiving applications have advanced searching capabilities). Instead of having to run the same search over and over across multiple systems, it would be great to have one place to conduct a federated search. However, the solutions landscape is in flux and organizations should start by exercising search and filtering capabilities on the most crucial content sources (e.g. email and file systems).

Organizations with small in-house legal staff should look to have an ECA interface that allows them to search and then conduct some level of actual analytics-driven review before paying for data processing. Features to look for include not only advanced search capabilities, but also concept searching, clustering, and perhaps visual analytics in addition to decision-support dashboards (e.g. reports on the estimated eDiscovery costs of a case based on the amount of data collected).

Both law firms and organizations with large in-house legal departments and/or organizations that are serial litigants should invest in advanced ECA tools that support iterative, transparent searching, visual analytics, concept searching, automated tagging, native document review, and document production. These types of tools are aimed at streamlining the actual legal review process and making the lawyers more efficient.

Many claim that there is no difference between traditional processing platforms and an “ECA application.” Both enable the user to index, search, and preview items - the minimum required to make strategic decisions early in the case lifecycle. Most full processing platforms process and explode a collection copy rather than the source ESI directly from the custodian’s location. This also requires a reasonably robust system, a technical user and can eat up a lot of storage overhead for the exploded (unpacking zips and email attachments) data and index. Up until recently, all of these systems had a volume-based (per GB) pricing model. This meant that a company had to spend time and money to collect ESI and have it rendered into a searchable form – all before management even had a good idea of the potential exposure. This is why corporate legal departments have not really embraced a full ECA process up to date. The typical time required for this out-sourced processing made high-speed reactive assessments almost impossible.

To become “ECA applications,” solutions must go further. Customers expect the software to enable search and data profiling either in place or at least without any additional expense or pre-processing. This shifts the user from a dedicated litigation support technician to an associate or

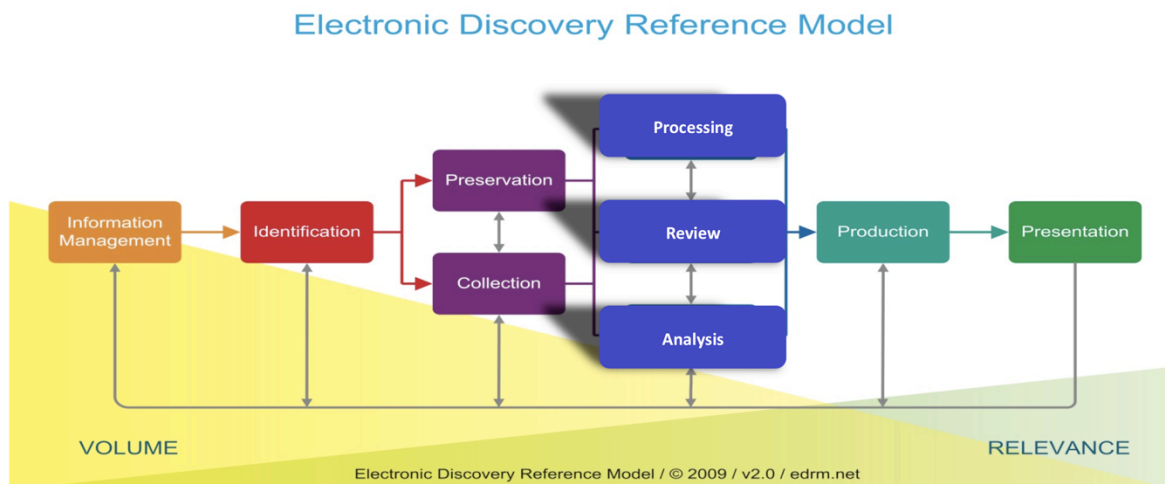
paralegal. The expectation is that the software can be pointed at a data source and be searched, profiled or explored without undue cost or effort. So the real differences are not in actual functionality, but in the graphical user interface (GUI) and workflow.

First-Generation And Next-Generation ECA

The market for “ECA solutions” is confusing because so many solution providers can, and will, claim to offer ECA. From simple search and filtering to advanced analytics, providers offer varying levels of legal decision support. For buyers, it’s important to understand requirements and how various solutions deliver features and functions to meet those requirements.

The first “ECA solutions” focused on taking collected data sets, processing the information, and then making it available for review. As such, these solutions focused on the process, analysis, and review modules of the EDRM, as seen in Figure 1 below.

Figure 1: First-Generation ECA



Processing, analysis, and review solutions can deliver real value. In looking at the vendor landscape, there are many features that are important for these components. The chart in Appendix A lists all the features that buyers should consider, as well as questions to ask of vendors in order to evaluate these features. Not all features are necessary of a good solution; each organization will be different. Buyers should be careful to define the features they require and prioritize them accordingly.

The clear trend, however, is for organizations to take more and more activity on the left side of the EDRM (information management, identification, preservation, and collection) in-house. As such, it is better to have a more integrated “ECA solution,” one that can identify potentially responsive information across multiple data sources, collect it, preserve it (either in-place or in a dedicated preservation repository), process it, run analytics, and present the information for review in a user interface that legal professionals can quickly grasp.

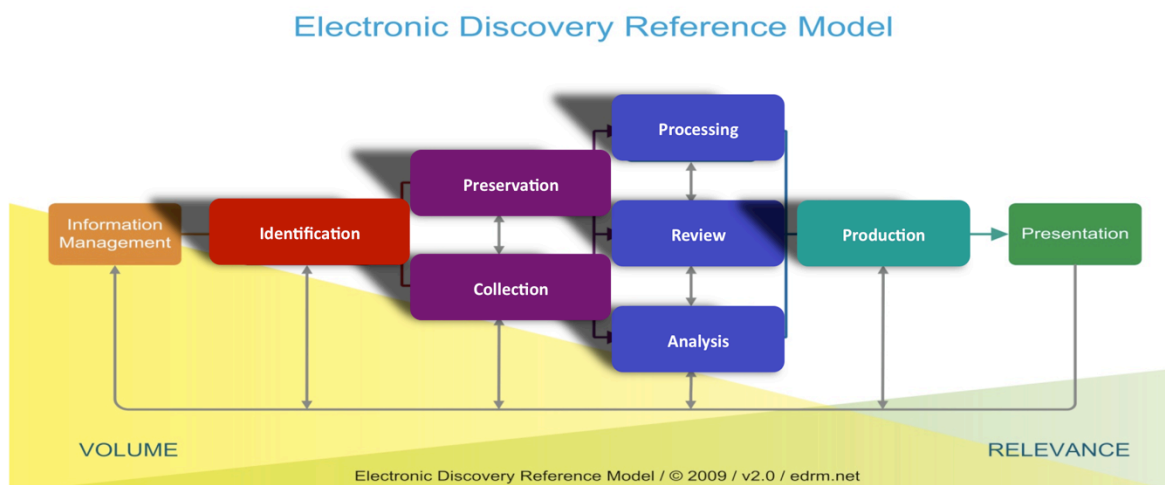
By adding functionality to identify, collect, and preserve, solution providers can deliver even more cost savings and risk mitigation. There is even the ability to conduct some level of analysis on information before it is collected. Some call this “very early case assessment” or “ECA in the wild.” No matter how it’s referred to, the end goal is having the organization know as much about potentially responsive information as early as possible.

A more integrated solution does not necessarily mean single-sourcing the ECA solution. As eDiscovery is an immature market, there are not many vendors that offer end-to-end solutions, and those that do often have modules that are weak links or simply a set of not-so-integrated products. The reality is that a full ECA solution may come from a number of vendors that partner to offer full functionality. It’s important to test that partnerships are real - that integrations are pre-built and proven as opposed to two vendors that have swapped logos for marketing purposes. Every time data is moved between applications, there is more risk introduced into the process. If integrations are not tested and proven, there is always the risk of potential spoliation.

The next-generation of ECA solutions adds not only identification, collection, and preservation features, but also production, as seen in Figure 2 below. While not always a necessity for corporations, production features allow organizations to quickly produce data to other parties as needed without having to send data to a law firm or out to a service provider for production services. Law firm ECA users will find that integrated production capabilities reduces the

complexity of transferring data from one tool to another just for the purpose of production. Such integration therefore saves time in addition to reducing risk.

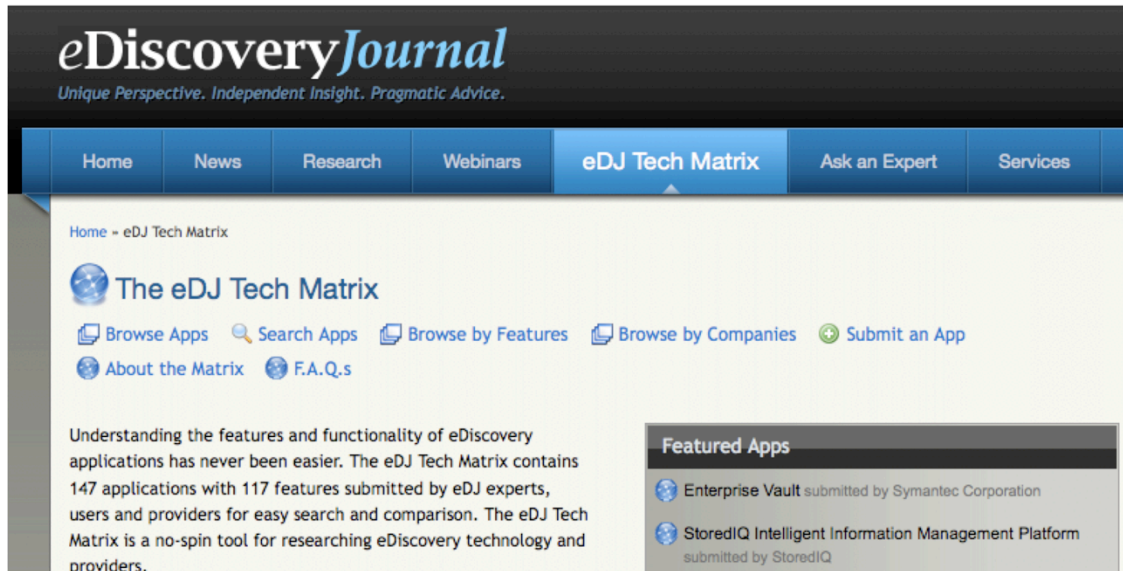
Figure 2: Next-Generation ECA



The chart in Appendix B lists all the features that buyers should consider in the identification, collection, and preservation and production arenas, as well as questions to ask of vendors in order to evaluate these features. Again, not all features are necessary of a good solution; each organization will be different. Buyers should be careful to define the features they require and prioritize them accordingly.

As mentioned, there are many software and service providers that offer ECA solutions. Instead of a static listing of providers in this report, eDiscoveryJournal provides a dynamic listing of applications and services that readers can browse through, search, and compare by feature. Simply to go to <http://ediscoveryjournal.com/the-edj-tech-matrix/> and use the features listed in Appendixes A & B to get the list of solutions that will meet requirements.

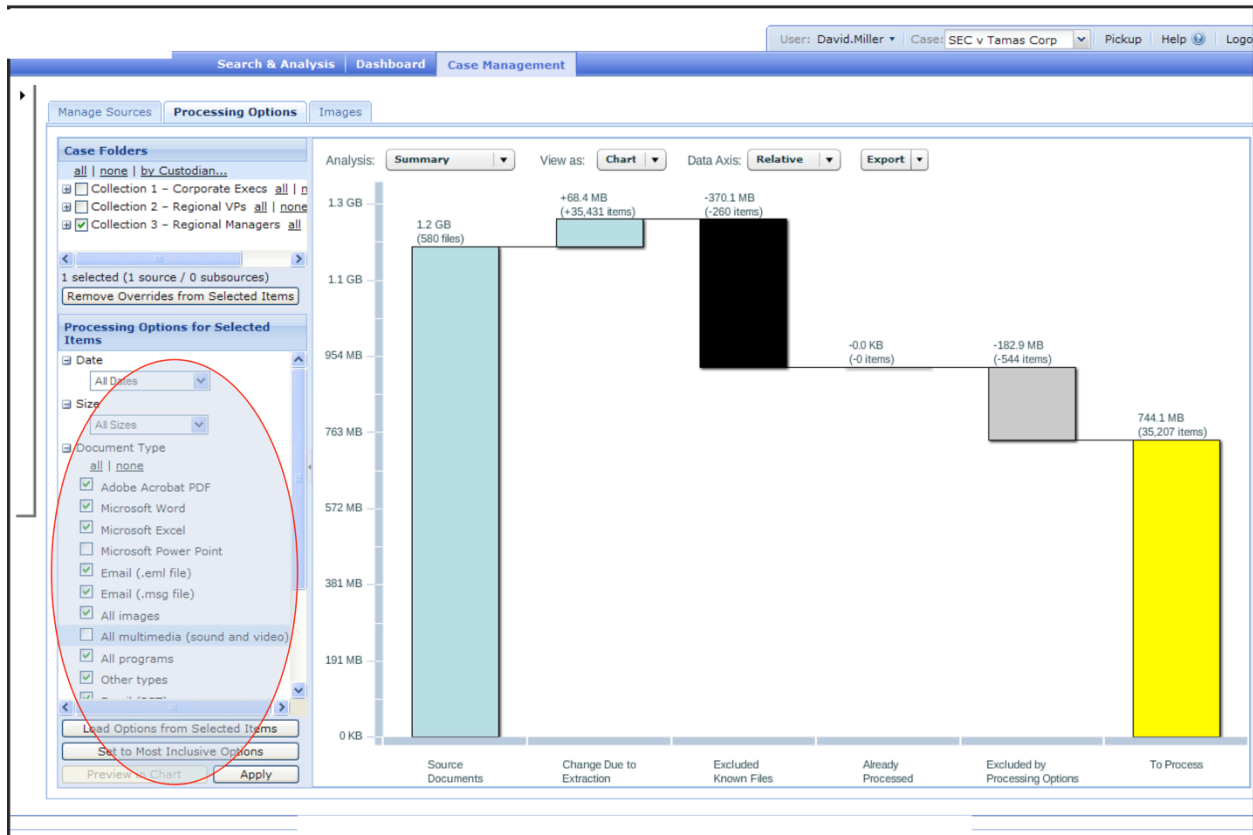
The eDiscoveryJournal Tech Matrix



WHAT DOES “ANALYTICS” REALLY MEAN?

Key to ECA is the Analysis component of the EDRM. The goal is to have applications analyze, prioritize, and even tag documents before a human looks at it. There are many different types of content analytics that can help ECA be a powerful tool.

- **Keyword culling** - simple and straightforward, keyword culling allows users to take negotiated keywords and phrases and eliminate any information not within the search results from the potentially responsive data set.
- **Metadata culling** - allows users to eliminate information from a potentially responsive data set based on information about the information (metadata). For example, any information outside of a certain date range can be eliminated, or data that does not match a certain content type (e.g. Word, Excel, .msg) can be excluded.



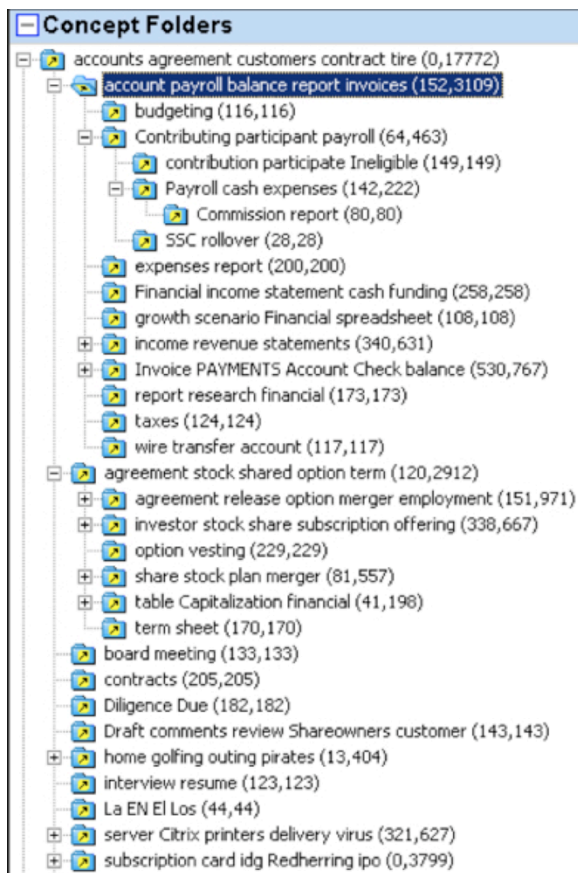
- **Faceted search** – ability to use metadata filters to further pivot on metadata fields; like getting rid of Amazon.com emails.

The screenshot displays the eDiscoveryJournal web interface. At the top, the user is logged in as 'superuser' and the case is identified as 'SEC v Tamas Corp'. The search results show 12,144 documents found. A sidebar on the left contains a 'Filter Search Results' section with various filters like 'By Folder', 'By Tag', and 'By Sender Domain'. The main area shows a list of documents with columns for 'Sender', 'Recipients', 'Date', and 'Score'. Each document entry includes a subject line and a brief preview of the content. A red circle highlights the 'Filter Search Results' sidebar.

Sender	Recipients	Date	Score
Todd Pickett (External)	Randy Gabrielson	02/01/2008 4:48 PM PST	
Todd Pickett (External)	Liz Chae	02/01/2008 4:47 PM PST	
Todd Pickett (External)	Steve Sher, Randy Gabrielson, Mike Simonsen	02/01/2008 4:45 PM PST	
Mike Simonsen (External)	Bob Bell	02/01/2008 4:38 PM PST	
Mike Simonsen (Internal)	Mike Tamas	01/24/2008 7:45 AM PST	
Steve Sher (Internal)	Randy Gabrielson, Mike Simonsen, Todd Pickett	01/24/2008 7:42 AM PST	
Steve Sher (Internal)	Randy Gabrielson	01/24/2008 7:41 AM PST	

- **Near deduplication** - unlike deduplication, which eliminates documents that are exact duplicates of others, near-deduplication allows users to eliminate or otherwise group documents that are materially similar, but not bit-level exact matches. Users must be careful, however, to repopulate near-dupes when producing the data set. A more common scenario would be to group near-dupes together so that the same reviewer is looking at all the similar documents, which can make the review process more efficient.
- **Discussion threading** – a feature used to keep conversations together. In the past, this has applied mostly to email – the most common collaboration mechanism. But, it is increasingly important for other collaborative tools such as bulletin boards, newsgroups, or social media. The solution aids the reviewers by visually grouping messages, typically in a hierarchy by topic. A set of messages grouped in this way is called a topic thread.

- **Concept clustering** - by grouping together potentially related documents, users have the ability to make the review process faster and/or more efficient. Concept clustering can take on various forms (as seen below) - grouping related documents into concept folders, creating heat maps of various concepts that users can click into - but is really about using machine algorithms to suggest groupings of content. It's also possible to use concepts to include or exclude whole groups of content from downstream activities. Beyond the ability to make review faster, it can allow users to optimize review resources. For example, high-cost associates could review the “hot” content while low-cost review resources look at the rest of the data set.



- **Predictive tagging / coding** - Predictive tagging, or coding, combines analytics with human review by taking sets of collected data (gathered by concept searching, phrase identification, keyword searching, metadata filters, etc) and having users review and code the documents for factors such as responsiveness, issue, or privilege. Applications can

then learn to tag similar documents based on the first set of human tagging. This process can reduce the total amount of documents reviewed and therefore has the potential for cost savings. In time, predictive tagging could ultimately become more common than traditional linear review. But, the legal industry is typically conservative and slow to adopt new technology and processes. If saving review costs is a priority, though, predictive coding is an option to investigate now. Note, it's important to conduct sampling of non-reviewed documents within data sets to ensure a good confidence interval of the process. Also, predictive tagging is not an all-or-nothing approach. Many solutions provide varying levels of confidence intervals and combine both machine and human intelligence so that organizations can determine their comfort with some level of computer prediction.

In addition to content analytics, process analytics drive effective ECA. Process analytics support decision-making by informing organizations about the cost to execute the eDiscovery process from collection through production. For example, companies can make better early decisions when they know metrics such as the average time it takes a specific firm to review a document, what a certain type of document typically costs to review, or how many documents are involved in an average FINRA investigation.

ECA - DIVERSE USERS AND A VARIETY OF USE-CASES

Part of what makes ECA confusing is the variety of ways in which ECA functionality can be applied. Different types of users may use ECA applications in dissimilar ways - one person's ECA might be filtering content based on keywords while another person's ECA is speeding investigation with hard-core visual analytics. First, let's examine the various types of end-users that require some level of ECA functionality (this is not meant to categorize these users, but rather list the many types that can be involved with ECA):

- Corporate IT
- Corporate Legal
- Corporate Litigation Support
- Corporate Records and Information Management
- Corporate Compliance
- Auditors / Corporate Investigators

- Law Firm IT
- Law Firm Legal
- Law Firm Litigation Support
- Project managers at vendors and law firms

Again, each of these users will likely be trying to accomplish something different with ECA. However, all of the uses of ECA support the five main goals that organizations seek to achieve via ECA:

- Cost reduction
- Better litigation win rates
- Strategic planning for matters based on merit
- Tactical planning for matters based on cost
- Litigation budget management

It's useful to cross-reference how various users leverage ECA in order to meet their goals, as shown below in Figure 3 below:

Figure 3 - How Users Leverage ECA to Achieve Goals

Goal	User	Function
Cost Reduction	Corporate IT, Corporate Records and Information Management	Reduce costs by defensibly disposing of data according to corporate retention policies Reduce cost of collection by centralizing collection interface to save time (or turn collection and preservation over to corporate legal)

Goal	User	Function
	Corporate Legal, Corporate Litigation Support	Reduce downstream processing and review costs by minimizing the universe of potentially responsive information via culling
	Law Firm Legal	Keep review costs down by prioritizing documents and assigning to the right level associates (better resource utilization)
	Law Firm Litigation Support	Reduce cost of review by culling information with advanced analytics
Better Litigation Win Rates	Corporate Legal, Corporate Litigation Support	Optimize decision-making (e.g. settling cases that can't be won) quickly with advanced analytics Quickly find the necessary information to win cases with advanced searches and prioritized review
	Law Firm Legal	Quickly find the necessary information to win cases with advanced searches and prioritized review
Strategic Planning For Matters Based on Merit	Corporate Legal, Corporate Litigation Support	Quickly determine if a matter has merit and is a winnable case
	Law Firm Legal, Law Firm Litigation Support	Quickly route prioritized documents to the right reviewers via advanced analytics, e.g. clustering

Goal	User	Function
Strategic Planning For Matters Based on Cost	Corporate Legal, Corporate Litigation Support	Quickly determine how much litigation will cost via early access to amount of potentially responsive information and prioritized review to make decisions based on the economics of the matter (e.g. settle for less than the cost of litigation)
Litigation Budget Optimization	Corporate Legal, Law Firm Legal	Minimize litigation budget by only pursuing winnable cases Minimize litigation budget by utilizing the lowest cost resources possible while putting high-cost resource on only the necessary documents

ECA ANECDOTES FROM THE TRENCHES

It's difficult to find public examples of best practices for ECA because most organizations are reluctant to talk openly about how they solve eDiscovery problems. Few companies want the world to know what their litigation costs are or how well their information is managed. Such public disclosure could put them at a disadvantage in future litigation. However, eDiscoveryJournal spoke with a financial services organization that serves as a great example of how ECA can provide benefits while also being an early step in a longer process of getting the full spectrum of eDiscovery under control.

Taking Corporate Control Of Litigation

One large financial services company could no longer bear the large review costs that resulted from turning over full collections to law firms for review. What this company wanted to do was "choose what, when, and how they do things vis-a-via litigation." Putting ECA in place allowed them to dictate how much information went to various law firms and gave the company the ability to make decisions quickly. What the legal team liked best was the measurable cost savings - on just one collection of 200,000 documents, the legal team was able to use ECA to

cull the reviewable set down to 200 documents. Of course, this was done after some training on iterative search and vetting of the non-responsive documents, but the takeaway is that this puts the decision-making power in the hands of corporate legal - they could review the documents themselves, or send the set out to a law firm knowing the overall review costs would be much lower. In this case, the benefit of ECA was greater than simple cost savings - the company now has control of the litigation process and is not constantly in a reactive mode and at the mercy of the law firms it works with.

Interestingly, while the legal team was the beneficiary of the ECA tool, IT was responsible for selecting the solution. Part of the goal was to reduce the tactical involvement of IT from the eDiscovery process, meaning that legal could run the application itself once IT had implemented it. Lessons learned from this company include:

While a single-appliance-based approach to ECA can be a good start to harnessing benefits, an enterprise-wide approach demands a more scalable solution – one that can handle the massive volumes of data in today’s organizations and process and analyze that data quickly. For a large organizations that views ECA as an enterprise business process, the single-appliance-based approach is not appropriate; rather, the lesson learned was to look for enterprise software with the associated scale and performance capabilities.

The user interface has to be legal friendly - the legal team will be using the application; if they don’t like the review interface, problems will ensue - it’s important to get legal’s buy-in early.

Architecture analysis and thorough testing are critical. Many tools have ECA interfaces slapped onto them, but are not one that legal teams could use due to the need to know scripting in the background. Nothing is more frustrating than having to call in IT resources repeatedly to write scripts.

Using Intelligence To Bring Efficiency To E-discovery Processes

Rare is the corporation that will talk publicly about eDiscovery efforts and/or processes.

Thankfully, David Steele, Senior Counsel, and Barbara Squires, Paralegal at Home Depot went on the record in a webinar titled “Home Depot Presents: The Five Habits to Create an Effective In-House E-Discovery Process,” documenting some effective in-house eDiscovery practices and

lessons learned.² For Home Depot, eDiscovery problems were becoming obvious. There was not a flexible, simple way to collect and preserve information in response to litigation. As a result, the company typically collected more than necessary. Over-collection is a problem, but one that is often overlooked. In Home Depot's case, the problem was exacerbated by certain cases with long timeframes. The company recognized the huge costs of over-collection when examining the cost of having data hosted with a third party service provider for review (and that money spent on over-collection really stings when the data is just sitting there as a case drags on).

The webinar focuses on five habits that can help create effective eDiscovery processes. Our analysis reveals what is really a three-part plan to implement the eDiscovery initiative:

- **Establish the present state.** For Home Depot, it was important to have an understanding of what the typical processes were vis-à-vis eDiscovery. Knowing the current processes allowed the company to capture case-level metrics, which in turn led to a holistic view of legal costs. It is this overall view of costs that became the impetus for change. Knowing that large sums of money are being wasted on unnecessary hosting costs allows for the business case creation.
- **Develop strategy for desired state.** In looking at the data about the current process, Home Depot was able to see that, while many cases were within individual case budgets, the overall litigation spend was very high. Having case-level metrics helped, too – in one case, only 5% of the collected data was produced to the opposing party. Home Depot was able to set a strategy: for small cases, try to handle most eDiscovery activities in house, e.g. quickly find the hot documents and make decisions; for large cases, make sure the data sent out for hosting and review was the smallest possible set of data. For Home Depot, knowing the current state allowed the company to set realistic goals for a desired state.
- **Get Legal and IT aligned.** eDiscoveryJournal has written extensively about process and technology ownership issues in eDiscovery. In fact, the inability to properly allocate roles and responsibilities within organizations is the primary reason for eDiscovery project failures. Home Depot recognized this issue early on and set up a cross-functional team made up of both legal and IT resources, and in addition had representation from

² From the Clearwell e-Discovery Webinar “Home Depot Presents: The Five Habits to Create an Effective In-House e-Discovery Process.” Available at http://www.clearwellsystems.com/e-discovery-resources/resource_dl_72.php

information security and forensics. Legal was able to sell IT on participation in the project because collection tools could be implemented to make IT’s job easier. Legal and IT worked together to define requirements. According to Home Depot, “Legal and IT became BFFs.” Home Depot created the the projects’ budget by forecasting the expected eDiscovery spend for the year (based on historical case facts) and then estimating the potential savings. As with any good project, there was an element of serendipity – the General Counsel had mentioned the potential savings to the CIO of the company; the CIO in turn mentioned it to a Senior VP of IT, to whom Legal was able to convince of the project’s merits. As a result, the eDiscovery project was funded by IT.

The Home Depot proves that it is possible to reap the benefits of a more proactive approach to eDiscovery – in this case, bringing efficiency to a process where costs had gotten too high and were avoidable. The savings Home Depot achieved are compelling – in just one case, there was a 95% cost reduction; a savings of almost \$500,000 from what the case would have cost if handled in the traditional manner:

	Before	After
Collected	200,000 docs	200,000 docs
Sent to Processing	200,000 docs	10,000 docs
Cost of Processing	\$75,000	\$3,750
Cost of Review	\$450,000	\$22,500
Access to Case Facts	Measured in weeks	Measured in hours
Duration	Measured in months	Measured in days

Not only did Home Depot see a significant cost savings, but the ancillary benefits were great, too. With access to case facts in hours instead of weeks, the legal team can quickly determine how to handle matters in the most efficient manner possible. Getting the duration of cases down from months to days allows the legal team to optimize resource usage and ensure that cost overruns won’t occur. The company took the right approach by getting legal and IT aligned

early on and identifying exactly how the project would be measured. Now, Home Depot can manage much of the eDiscovery process in-house and when it is necessary to send data out for processing and review, the company can know that it is a small subset of what it would have sent out before this project.

MOVING FORWARD WITH ECA

Clearly, ECA initiatives can provide multiple benefits, but implementing ECA will be different for every organization. In general, when moving forward with ECA, organizations should should:

- **Allocate legal and IT responsibilities correctly at the outset.** With ECA, legal is the business user of the application. As such, the legal team should set the requirements for usability and features and functions of a solution. IT should be involved when ECA solutions are to be deployed within an organizations IT environment or when the solution will connect to various data sources within the IT environment. The market of ECA consumers is too fragmented to simply state that legal should set requirements and IT should make the purchase. That philosophy only applies when referring to a software purchase. If a legal team wants to use an EDD service provider for ECA functionality, it should have the right to purchase that. IT and legal can then work together to ensure an efficient way to identify and collect data to be given to the EDD service provider. IT and/or risk management should also work to ensure that the service provider complies with expected corporate network security requirements.
- **Consider current volume of matter and matter data.** There is a lot of hype around scale and performance of ECA solutions. For many organizations, however, a small ECA appliance will handle the current volume of information involved in litigation. As larger enterprises roll out ECA tools and integrate the identification and collection initiatives, speed and scale will become a more important priority and IT teams should plan as such.
- **Understand that the solutions market is immature.** Just because a couple of vendors can claim to have applications that address the full spectrum of the EDRM does not mean that all functionality is best-of-breed. The financial services company referenced earlier does not plan to buy identification, preservation, and collection functionality from its ECA vendor despite the fact that they are happy with the ECA functionality and the vendor offers a product for identification, preservation, and collection. The company is comfortable with sourcing from multiple vendors to build the best eDiscovery

capabilities that it can. Be sure to purchase what meets your needs now and look closely at the vendor's plans for the future to be sure it can evolve with you.

- **Insist on a proof-of-concept.** Vendors tout proprietary algorithms and patented content analytics - and those things are great. But, conducting a patent analysis is hard work and it's difficult to compare one vendor's patent versus another's - and it's likely a waste of time. Instead, test 2-3 solutions from a short list. A POC not only allows you to get a feel for the user interface and learn if legal will be able to use the application, but it also allows you to see how the application performs on your data, which can let you compare applications in an apples-to-apples fashion. It is important to run your data through potential software rather than rely on preloaded Enron email for your evaluation. Your data has unique properties; use and test it before you buy. The same thinking applies whether considering software or services; insist on a POC with an EDD service provider before investing fully. And, for organizations just beginning to test the water, consider starting with a small deployment while making sure it has the potential to scale out to enterprise-grade ECA.
- **Measure ROI on more than just cost savings.** Yes, one of the primary benefits of ECA is cost reduction, but it is wise to begin measuring all eDiscovery initiatives on how they impact the lifecycle of matters. The efficiencies gained in collecting information, for example, have benefits that go way beyond reduced cost - the IT time not wasted on reactive collection is more time available for innovative projects that drive revenue for companies.
- **Double-check that existing tools will provide functionality that fits ECA requirements.** Not all tools were built to provide analytics and support fast decision-making. Ensure that providers of related tools such as archives and preservation repositories are able to provide, or will soon be able to provide, the right set of features and functions to support ECA activities.

APPENDIX A - PROCESSING, ANALYSIS, AND REVIEW FEATURES

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Annotations	Comments, annotations and highlighting that is searchable	Do you provide the ability to annotate ESI including comments and highlighting that is searchable? Please describe.
Audio Search	Ability to search by conversion to text or direct phonetic search of audio/video format ESI	Are you able to search audio and video format ESI? If so, please describe how. Is this capability fully integrated or does audio / video ESI need to be processed separately? Is this functionality native or provided via partner / OEM?
Audit Trail	System securely logs and reports on actions	Do you provide an audit trail that securely logs and reports on system actions? Please describe.
Auto Redaction	Place redactions on native ESI by rule	Do you support automated redaction of native or image ESI files based on rules? Please describe.
Autocode	Ability to generate objective or subjective fielded data from ESI through automated process, also called Autocategorization	Are you able to provide auto-categorization of ESI? Are you able to provide suggested tags for ESI? How is this functionality implemented? Is it native or provided via partner / OEM?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Automated Review	Automated relevance or issue designation upon a collection without individual manual viewing of every item. This includes using training sets, expansion by cluster or other iterative pattern systems.	Do you provide automate issue designation without reviewer intervention? Do you combine this with machine learning? How is this functionality implemented? Is it native or provided via partner / OEM?
Case Management	Workflow, timelines and other tools to track and report on case status, productivity and overall progress	Do your provide the ability to manage individual cases via workflow and tracking capabilities?
Chain of Custody	Automated creation of Chain of Custody documentation that tracks collection and transmission of ESI.	How do you monitor the chain of custody for all ESI within a given matter?
Chronological Analysis	Analysis of ESI chronology	Are you able to conduct chronological analysis of ESI? Please describe.
Communication De-Duplication	Metadata level deduplication of email, IM and other communication format ESI	With respect to communication-based ESI (e.g. email , IM), do you offer metadata level de-duplication? Please describe.

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Concept Search	Advanced index that extracts concepts	Do you offer concept search (advanced index that extracts concepts)? Is this capability native or provided via partner / OEM?
Concept Clustering	Present documents in a manner that visually segments them based on the concepts they are potentially related to	Do you provided the ability to visually present documents sorted by concepts, e.g. folders, heat map?
Conceptual Review	Non-linear review by concept or other clusters for increased efficiency	Do you provide the ability to review ESI based on concepts either by clustering or foldering ESI into concept-based organizational structure? Please describe and include a screenshot of the UI.
Conditional Tagging	Create customized tag or designation rules that exclude tag combinations or force additional tags.	Do you provide the ability to create customized tag or designation rules that exclude tag combinations or force additional tags? Please describe.
Content Culling	Removal and suppression of ESI based on search terms and item content rules	Do you offer culling of content based on search terms and/or item content rules? Is this capability native or provided via partner / OEM?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Content Flagging	Rule based flagging, tagging and categorization to aid review and processing	Do you provide the ability to manually tag ESI (e.g. privileged, responsive)? Are you able to provide suggested or automated tagging of ESI? Do you provide any kind of automated categorization of ESI for the purpose of managing the review process? Are these capabilities native or provided via partner / OEM?
Conversation and thread Analysis	Analysis of communications to establish conversational relationships in email as well as beyond email threading	Are you able to conduct email thread conversation analysis? Are you able to establish conversational relationships beyond email thread analysis? Please describe.
Customizable Reports	Ability for users and administrators to create custom reports without scripting or leaving the user interface.	Do you provide the ability for users to create customized reports without scripting and without leaving the user interface? Please describe.

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Decision Documentation	Ability to provide reports that auto-document the audit trail on search decisions made during searches and refinements in order to provide documentation for how iterative searches were conducted.	Do you provide reports that auto-document decisions made during the search and refinement process? If so, how are such reports configured and executed? Can you provide copies of example reports?
De-NIST Culling	Removal of known hash sets like the National Institute of Science and Technology known file list	Are you able to remove known hash sets or file sets based on lists like the National Institute of Science and Technology known file list?
Decryption	Decrypts files either with known master and user keys or actually cracking encrypted files.	Do you have the capability to decrypt files? How is this enabled?
Domain Analysis	Extraction of domain level communication patterns from ESI	Are you able to conduct analysis of ESI whereby domain level communication patterns are extracted from ESI? Please describe.

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Dynamic Folders	ESI folders or groups dynamically populated based on rules, text content and field information, such as by custodian, date, custodian and date, file extension, inbox, or file path.	Do you provide foldering, the ability to dynamically create groupings of ESI based on rules, content, or metadata? Please describe. Is this capability provided natively or via partner / OEM?
Electronic File Numbering	Assign bates numbers and rename or embed metadata tag for native ESI productions	Can you assign bates numbers and rename or embed metadata tag for native ESI productions? Can you assign bates number in a gap-free manner? Is this functionality native or provided via partner / OEM?
Electronic Image Numbering	Assign bates numbers and brand image sets. Bates numbering (also known as Bates stamping or Bates coding) is used in the legal, medical, and business fields to place identifying numbers and/or date/time-marks on images and documents as they are scanned or	Can you assign bates numbers and brand them into image sets? Can you assign bates number in a gap-free manner? Is this functionality native or provided via partner / OEM?
Email Thread De-Duplication	Identification and suppression of prior duplicative emails in threads	Are you able to de-duplicate email based on thread analysis?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
ESI Sampling	System for random sampling that can be used to establish ESI characteristics or in quality control or assurance.	Do you provide the ability to randomly sample ESI to establish characteristics or quality control? How is this enabled?
Email collection	Ability to connect to and collect ESI/email from a live email server or a file such as EDB (Exchange database)	Are you able to search live email servers or mail databases in order to collect email content? How is this functionality executed?
Faceted Search	Ability to use metadata filters to further pivot on metadata fields; like getting rid of Amazon.com emails	Do you provide metadata filters that in turn allow users to visualize search results based on metadata fields, e.g. bucketing all emails from a specific domain into one folder for bulk actions?
File Near De-Duplication	Identification and suppression of near duplicate files	Do you provide near de-duplication capabilities? Is near de-dupe native or provided via partner / OEM?
File Type Identification	Identifies file type by content, not just by file extension.	Are you able to identify file types based on content (not just the file extension)?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Forecasting Metrics	Generate cost, time or volume forecasting reports or dashboard to create or reassess estimates according to new project assumptions.	Do you provide a dashboard view of cost, time, and/or volume forecasting reports to create or reassess estimate based on project assumptions or real-time status? If so, how is this enabled?
Foreign Language Support	Application is capable of processing or searching foreign languages and uncommon character sets such as double-byte characters.	Do you provide foreign language support such that you can process and search foreign languages and uncommon character sets such as double-by characters?
Fuzzy Logic Search	Partial word recognition to compensate for OCR and spelling issues	Do you provide fuzzy logic search for partial word recognition to compensate for OCR and spelling issues? Is this functionality native or provided via partner / OEM?
Generate Load Files	Creation of load files for specific applications to aid the transmission of ESI	Are you able to generate load files for transmitting ESI to other applications? Please list all load formats you support.
Hash De-Duplication	File and container level deduplication by Hash value	Do you offer de-duplication? If so, what methods of de-duplication do you use?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Image Conversion	Convert ESI into a fixed image format	Are you able to convert ESI into a fixed image format? If so, how this accomplished? Is the capability native or via partner / OEM?
Indexed Search	Creation of an index for search	Do you create an index of ESI for the purpose of searching it? Is the indexing technology native or via partner / OEM? How is the index created and where is it stored? What is the average ratio of index size to ESI source size?
Inventory	Generates a file inventory or catalogue of the targeted data sources	Do you generate an inventory of all targeted data sources? How is this assembled?
Linguistic Analysis	Extraction of linguistic relationships from ESI	Are you able to conduct linguistic analysis whereby linguistic relationships are extracted from ESI? Please describe.
Machine Text Translation	Automated translation of foreign language files and communications into English.	Do you support foreign languages files and translate them to English? How is this accomplished?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Media Restoration	Restores ESI from archival or backup media to active file system	Do you provide tape / backup media restoration capabilities? If so, are these capabilities native or via partner / OEM?
Metadata Culling	Removal and suppression of ESI by type, date and other metadata properties	Do you offer culling of ESI based metadata properties like type or date? Please describe. Is this capability native or provided via partner / OEM?
Metadata Extraction	Extracts file and system metadata from files and communication systems.	Are you able to extract file and system metadata from all ESI data sources? Please list all sources.
Non-text File Detection	Detects image, audio, animation and other non-textual file types and attachments. May have workflow capabilities	Are you able to automatically detect and report on non-text file types and attachments? How do you deal with detected non-text file types?
OCR	OCR, HCR, ICR of image format ESI to extract text	Are you able to index ESI in image format to extract text? Is OCR capability fully integrated into your solution or is OCR data processed separately? Is OCR functionality native or provided via partner / OEM?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Parallel Review	Supports complex parallel simultaneous reviews on a single matter/collection. This can include different tagging rules, groupings and filters. The core function is the ability to host overlapping independent review workflows on the same matter.	Do you support complex parallel simultaneous reviews on a single matter, including different tagging rules, groupings, and filters? Please describe.
Password Removal	Removes password protection from ESI either from a list of known password or by brute force password cracking.	Are you able to remove passwords from ESI? Please explain any methodologies used.
Performance Metrics Analysis	Automated tracking and analysis of metrics associated with discovery steps and phases. The results of the metrics analysis are presented via summary dashboard or reports to enable iterative feedback to optimize and correct processes.	Do you provide automated tracking and analysis of metrics associated with eDiscovery? Please provide a list of reports and screenshot of the UI.

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Privacy Rights Management and Compliance	An ability to support discovery workflow and processes while managing user privacy issues and in compliance with foreign privacy laws such as the European Union.	Are you able to support eDiscovery actions like collection and legal hold while managing user privacy issues like the privacy laws that exist in the European Union? Please describe how.
Privilege Log	Create and manage privilege logs within application from ESI metadata. Logs can also be edited and linked back to the privilege documents.	Are you able to create and manage privilege logs from ESI metadata? Can the logs be edited and linked back to the privileged documents? Please describe.
Redact Images	Place redactions on images for branding	Do you support redactions on image files? Is this capability native or provided via partner / OEM?
Redact Native ESI	Place redactions on native ESI and then output redacted version	Do you support redactions on native ESI files? How does this capability work? Is this capability native or provided via partner / OEM?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Remote Hosted Administration	Hosted platform that supports remote administration by the client to perform most functions without reliance upon or notification to the provider to perform actions behind the firewall.	Do you provide clients' the ability to perform most or all eDiscovery functions with remote administration behind their firewalls or in other locations? Please describe how this is enabled?
Remote Review	System supports secured remote review	Do you support secured remote review? Please describe.
Repopulate Duplicates	Repopulate all near and full duplicates on production	Are you able to repopulate all near and full duplicates on production? Please describe.
Review Images	Review and designate images connected to database	Do you provide the ability to review images (e.g. TIFF, PDF)? Please provide a screenshot of the user interface.
Review Management	Enables assignment, tracking, QA/QC and other review management functions	Do you provide review management including assignment, tracking, and QA/QC? Please describe.

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Review Native ESI	Review and designate ESI in native or original format	Do you provide the ability to review ESI in its native format (e.g. spreadsheet)? How is this accomplished? How do you ensure that metadata for the document does not change? Please provide a screenshot of the user interface.
Role Management	Manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level.	Do you provide the ability to manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level? Please describe?
Rule-based Dynamic review management	Automated, dynamic management of review batches based on rules. Rules Based Review Batches, Dynamic Review Batches, Query Based Review Batches, Cascading Review Queues	Do you provide automation of review based on rules and/or analytics? Are you able to provide dynamic review batches? If so, how is this functionality enabled?
Scan Images	Scan physical paper into image format	Are you able to scan paper into an indexable image format? If so, is this capability native or provided via partner / OEM?

Processing, Analysis, and Review Features		
Feature	Feature Description	RFP Questions
Secured Subset Review	Ability to give secured access to subsets of collection for in camera, opposing or expert review	Do you provide the ability to give secured access to subsets of collection for in camera, opposing or expert review? Please describe.
Skin Tone Analysis	Detect and analyze skin tone patterns within images to find potential pornography and inappropriate content.	Are you able to conduct skin tone analysis to detect patterns within images to find potential inappropriate content (e.g. pornography)? Please describe.
Social Networking	Analysis of communication metrics to extract social network information	Are you able to conduct analysis of ESI to extract social network information (e.g. who communicates with whom)? Please describe.
Text Extraction	Extract raw or formatted text from ESI	Are you able to extract raw or formatted text from ESI? How is this accomplished? Is the functionality native or provided via partner / OEM?
Unpack Container Files	Unpack, explode or extract all container files such as ZIP, CAB and email containers such as NSF and PST files along with extraction of attachments from emails.	Are you able to process container files like ZIP, CAB, NSF, and PST (along with extracting attachments from emails)? Is this capability native or via partner / OEM?

APPENDIX B - IDENTIFICATION, COLLECTION, PRESERVATION, AND PRODUCTION FEATURES

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Audio Search	Ability to search by conversion to text or direct phonetic search of audio/video format ESI	Are you able to search audio and video format ESI? If so, please describe how. Is this capability fully integrated or does audio / video ESI need to be processed separately? Is this functionality native or provided via partner / OEM?
Audit Trail	System securely logs and reports on actions	Do you provide an audit trail that securely logs and reports on system actions? Please describe.
Autocode	Ability to generate objective or subjective fielded data from ESI through automated process, also called Autocategorization	Are you able to provide auto-categorization of ESI? Are you able to provide suggested tags for ESI? How is this functionality implemented? Is it native or provided via partner / OEM?
Chain of Custody	Automated creation of Chain of Custody documentation that tracks collection and transmission of ESI.	How do you monitor the chain of custody for all ESI within a given matter?

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Chronological Analysis	Analysis of ESI chronology	Are you able to conduct chronological analysis of ESI? Please describe.
Cloud Collection Support	Ability to collect from externally hosted repositories of content such as SaaS email archives.	Does the solution collect from third-party, hosted repositories? If so, list all systems? For each system, please document both the API connector and any collection methodologies (both for file content and metadata), as well as data movement protocols.
Communication De-Duplication	Metadata level deduplication of email, IM and other communication format ESI	With respect to communication-based ESI (e.g. email , IM), do you offer metadata level de-duplication? Please describe.
Concept Search	Advanced index that extracts concepts	Do you offer concept search (advanced index that extracts concepts)? Is this capability native or provided via partner / OEM?
Content Culling	Removal and suppression of ESI based on search terms and item content rules	Do you offer culling of content based on search terms and/or item content rules? Is this capability native or provided via partner / OEM?

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Crawl Search	Use of GREP or other search that actively checks files without creation of index. This is a point in time search.	Are you able to actively search files without creation of an index? How is this accomplished?
Custodian Collection	Web based or enterprise application that assists Custodians in designating and collecting ESI.	Do you allow for custodian-led collection? How is this process enabled? Please describe how this is tracked and the components of the interface for custodians. Is this capability native or via partner / OEM?
Customizable Reports	Ability for users and administrators to create custom reports without scripting or leaving the user interface.	Do you provide the ability for users to create customized reports without scripting and without leaving the user interface? Please describe.
De-NIST Culling	Removal of known hash sets like the National Institute of Science and Technology known file list	Are you able to remove known hash sets or file sets based on lists like the National Institute of Science and Technology known file list?

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Desktop Search	Search across live network to remote user desktop and laptop devices	Do you provide the ability to search across across live network to remote desktops and laptops? Is this capability native or provided via partner / OEM? Please describe all desktop search and collection capabilities.
Domain Analysis	Extraction of domain level communication patterns from ESI	Are you able to conduct analysis of ESI whereby domain level communication patterns are extracted from ESI? Please describe.
ESI Sampling	System for random sampling that can be used to establish ESI characteristics or in quality control or assurance.	Do you provide the ability to randomly sample ESI to establish characteristics or quality control? How is this enabled?
Email collection	Ability to connect to and collect ESI/email from a live email server or a file such as EDB (Exchange database)	Are you able to search live email servers or mail databases in order to collect email content? How is this functionality executed?
File Type Identification	Identifies file type by content, not just by file extension.	Are you able to identify file types based on content (not just the file extension)?

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Foreign Language Support	Application is capable of processing or searching foreign languages and uncommon character sets such as double-byte characters.	Do you provide foreign language support such that you can process and search foreign languages and uncommon character sets such as double-by characters?
Forensic Image	Physical level forensic copy of the media that includes all slack space, deleted files, fragments and other content.	Are you able to create forensic images of ESI sources? Please indicate what sources. How is the forensic image created? Does it include all slack space, deleted files, fragments, and other content on the ESI source? Is this capability native or via partner / OEM?
Forensically Sound Copy	Verified content, file metadata and system metadata by Hash and properties check. Many applications also generate a Chain of Custody log	How do you monitor and maintain a forensically sound chain of custody for all content, file metadata, and system metadata? Please describe the methodology in detail.
Generate Load Files	Creation of load files for specific applications to aid the transmission of ESI	Are you able to generate load files for transmitting ESI to other applications? Please list all load formats you support.

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Hash De-Duplication	File and container level deduplication by Hash value	Do you offer de-duplication? If so, what methods of de-duplication do you use?
Indexed Search	Creation of an index for search	Do you create an index of ESI for the purpose of searching it? Is the indexing technology native or via partner / OEM? How is the index created and where is it stored? What is the average ratio of index size to ESI source size?
Inventory	Generates a file inventory or catalogue of the targeted data sources	Do you generate an inventory of all targeted data sources? How is this assembled?
IT Source Map	Master list of IT assets and ESI Repositories	How do you create a master list of all IT assets and ESI repositories?
Legal Hold Notices	Create, transmit and track Legal Hold notices	Do you create, transmit, and track legal hold notices? If so, how? Is this capability native or provided via partner / OEM?
Linguistic Analysis	Extraction of linguistic relationships from ESI	Are you able to conduct linguistic analysis whereby linguistic relationships are extracted from ESI? Please describe.

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Machine Text Translation	Automated translation of foreign language files and communications into English.	Do you support foreign languages files and translate them to English? How is this accomplished?
Media Restoration	Restores ESI from archival or backup media to active file system	Do you provide tape / backup media restoration capabilities? If so, are these capabilities native or via partner / OEM?
Metadata Culling	Removal and suppression of ESI by type, date and other metadata properties	Do you offer culling of ESI based metadata properties like type or date? Please describe. Is this capability native or provided via partner / OEM?
Network Search	Search across live enterprise network data sources	Do you provide the ability to search across live network data sources? Is this capability native or provided via partner / OEM? Please describe all search capabilities.
Non-Rights Based Preservation In Place	System does not modify file rights to preserve ESI, but preserves ESI in its original location by use of other means such as agents, servlets, backup system, etc.	Are you able to put ESI on legal hold in its original location? If so, how is this accomplished? Please describe the repositories that you have active integrations to for in-place legal hold.

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Non-text File Detection	Detects image, audio, animation and other non-textual file types and attachments. May have workflow capabilities	Are you able to automatically detect and report on non-text file types and attachments? How do you deal with detected non-text file types?
Organization Map	Relational map of organizational entities and reporting structure; can be based off of organization charts or dynamically controlled by HR system	Can you create an organizational map of users / custodians and their reporting structures? How is this accomplished? If via HR systems, please list all systems supported and describe the level of integration.
Performance Metrics Analysis	Automated tracking and analysis of metrics associated with discovery steps and phases. The results of the metrics analysis are presented via summary dashboard or reports to enable iterative feedback to optimize and correct processes.	Do you provide automated tracking and analysis of metrics associated with eDiscovery? Please provide a list of reports and screenshot of the UI.
Preservation Repository	System designed to preserve ESI by holding and tracking a copy	Do you have the ability to copy ESI into a preservation repository? If so, is the repository your own? What repositories can you copy ESI to for legal hold?

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Privacy Rights Management and Compliance	"An ability to support discovery workflow and processes while managing user privacy issues and in compliance with foreign privacy laws such as the European Union.	Are you able to support eDiscovery actions like collection and legal hold while managing user privacy issues like the privacy laws that exist in the European Union? Please describe how.
Response Workflow	Tracking and workflow system that documents event response actions	What kind of tracking and workflow systems to document event response actions exists? Is the technology organic or sourced via OEM? Please describe.
Repository Support	Ability to collect from all repositories of content such as email, archives, file systems, ECM systems, collaboration systems, file systems, databases	What repositories do you support collecting ESI from? For each system, please document both the API connector and any collection methodologies (both for file content and metadata).
Rights Based Preservation In Place	System modifies file rights to preserve ESI in original location by search or active rule policy	Are you able to put ESI on legal hold in its original location? If so, how is this accomplished? Please describe the repositories that you have active integrations to for in-place legal hold.

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Role Management	Manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level.	Do you provide the ability to manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level? Please describe?
SharePoint Collection	Ability to collect files and other content (e.g. metadata, blogs, wikis, tasks, workflow associations, UI context, etc) from SharePoint.	Do you provide the ability to collect content from SharePoint? For content other than document libraries, please describe how collection is enabled, chain-of-custody is managed, and how the data is presented for review. Is this functionality native or partner-provided?
Skin Tone Analysis	Detect and analyze skin tone patterns within images to find potential pornography and inappropriate content.	Are you able to conduct skin tone analysis to detect patterns within images to find potential inappropriate content (e.g. pornography)? Please describe.
Social Networking	Analysis of communication metrics to extract social network information	Are you able to conduct analysis of ESI to extract social network information (e.g. who communicates with whom)? Please describe.

Identification, Collection, and Preservation Features		
Feature	Feature Description	RFP Questions
Stealth Hold	Legal Hold of documents without impacting end users or informing them of the hold.	Do you provide the ability to put ESI on legal hold without informing the custodian or impacting the custodian? If so, please describe how this is accomplished.
Stealth Investigation	Ability to search, browse and collect from custodians's data sources without alerting target custodian or impacting their system adversely.	Do you provide the ability to search and collection ESI from custodian's data sources without impacting the custodian or alerting the custodian? If so, please describe how this is accomplished.
User Map	Master list of users; usually from Active Directory	How do you create a master list of users/custodians? Can you integrate with Active Directory and other LDAP systems? Please list all.

Production Features		
Feature	Description	RFP Questions
Audit Trail	System securely logs and reports on actions	Do you provide an audit trail that securely logs and reports on system actions? Please describe.

Production Features		
Feature	Description	RFP Questions
Auto Redaction	Place redactions on native ESI by rule	Do you support automated redaction of native or image ESI files based on rules? Please describe.
Bulk Printing	Large scale blowback of branded images	Do you provide for bulk printing of ESI sets including large-scale blowback of branded images? Please describe.
Chain of Custody	Automated creation of Chain of Custody documentation that tracks collection and transmission of ESI.	How do you monitor the chain of custody for all ESI within a given matter?
Customizable Reports	Ability for users and administrators to create custom reports without scripting or leaving the user interface.	Do you provide the ability for users to create customized reports without scripting and without leaving the user interface? Please describe.
Disposal Certification	Deletion of ESI with expiry/destruction certificates	Do you provide disposal certification including deletion of ESI with expiry/destruction certificates? Please describe.

Production Features		
Feature	Description	RFP Questions
Electronic File Numbering	Assign bates numbers and rename or embed metadata tag for native ESI productions	Can you assign bates numbers and rename or embed metadata tag for native ESI productions? Can you assign bates number in a gap-free manner? Is this functionality native or provided via partner / OEM?
Electronic Image Numbering	Assign bates numbers and brand image sets. Bates numbering (also known as Bates stamping or Bates coding) is used in the legal, medical, and business fields to place identifying numbers and/or date/time-marks on images and documents as they are scanned or	Can you assign bates numbers and brand them into image sets? Can you assign bates number in a gap-free manner? Is this functionality native or provided via partner / OEM?
ESI Sampling	System for random sampling that can be used to establish ESI characteristics or in quality control or assurance.	Do you provide the ability to randomly sample ESI to establish characteristics or quality control? How is this enabled?
Exception Workflow	Batch level workflow that allows reprocessing, insertion of placeholders and exception reports	Do you offer workflow capabilities to allow for reprocessing, insertion of placeholders, and exception reports? Please describe.

Production Features		
Feature	Description	RFP Questions
Forecasting Metrics	Generate cost, time or volume forecasting reports or dashboard to create or reassess estimates according to new project assumptions.	Do you provide a dashboard view of cost, time, and/or volume forecasting reports to create or reassess estimate based on project assumptions or real-time status? If so, how is this enabled?
Generate Load Files	Creation of load files for specific applications to aid the transmission of ESI	Are you able to generate load files for transmitting ESI to other applications? Please list all load formats you support.
Image Conversion	Convert ESI into a fixed image format	Are you able to convert ESI into a fixed image format? If so, how this accomplished? Is the capability native or via partner / OEM?
Multiple Party Productions	Supports tracking and coordination of multiparty/matter overlapping productions	Do you support tracking and coordination of multiparty/matter overlapping productions? Please describe.

Production Features		
Feature	Description	RFP Questions
Performance Metrics Analysis	Automated tracking and analysis of metrics associated with discovery steps and phases. The results of the metrics analysis are presented via summary dashboard or reports to enable iterative feedback to optimize and correct processes.	Do you provide automated tracking and analysis of metrics associated with eDiscovery? Please provide a list of repots and screenshot of the UI.
Privacy Rights Management and Compliance	"An ability to support discovery workflow and processes while managing user privacy issues and in compliance with foreign privacy laws such as the European Union.	Are you able to support eDiscovery actions like collection and legal hold while managing user privacy issues like the privacy laws that exist in the European Union? Please describe how.
Redact Images	Place redactions on images for branding	Do you support redactions on image files? Is this capability native or provided via partner / OEM?
Redact Native ESI	Place redactions on native ESI and then output redacted version	Do you support redactions on native ESI files? How does this capability work? Is this capability native or provided via partner / OEM?

Production Features		
Feature	Description	RFP Questions
Remote Hosted Administration	Hosted platform that supports remote administration by the client to perform most functions without reliance upon or notification to the provider to perform actions behind the firewall.	Do you provide clients' the ability to perform most or all eDiscovery functions with remote administration behind their firewalls or in other locations? Please describe how this is enabled?
Repopulate Duplicates	Repopulate all near and full duplicates on production	Are you able to repopulate all near and full duplicates on production? Please describe.
Role Management	Manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level.	Do you provide the ability to manage and customize multiple security and user roles for various tasks and responsibilities on a matter and application level? Please describe?
Text Extraction	Extract raw or formatted text from ESI	Are you able to extract raw or formatted text from ESI? How is this accomplished? Is the functionality native or provided via partner / OEM?

© 2010, eDiscoveryJournal, LLC. All rights reserved. Customers that bought this report may make one attributed copy or slide of each figure contained herein. Additional reproduction is strictly prohibited. For additional reproduction rights and usage information, go to www.ediscoveryjournal.com. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.