

## 5 Critical Factors to In-House eDiscovery Success

Clearwell eDiscovery Platform 5.0

August 2009



eDiscovery for civil litigation is a fact of life in the enterprise. It is also an expensive fact: U.S. corporations spend nearly \$2 billion a year for eDiscovery-related tasks.

eDiscovery was not always such a thorn in the corporate side. Back in the days of manageable record volumes, attorneys were used to throwing armies of contract attorneys to review some letter boxes and a few gigabytes of electronic data. But now it's a whole new ballgame. The same armies are out there but now they must analyze terabytes of electronic data, shrink that data to manageable review levels, produce millions of pages – oh, and they must do it fast. Many attorneys try to do what they always did by pleading the burdensome costs of eDiscovery to the court. That used to work and still does in some cases, but judges are less and less inclined to cut the complaining party some slack.

This perfect storm of eDiscovery requirements is not going away and the enterprise is bearing heavy costs for it. If the corporation has any hope of reining in steep eDiscovery costs they must find powerful products to help them do it. However, navigating the marketing hype to understand the real capabilities of “eDiscovery” solutions can be difficult given the vendor land grab. Buyers must not only understand a product's breadth (how many eDiscovery stages it supports) but also how deeply it supports them. Without necessary depth, some eDiscovery products will be unable to control real eDiscovery costs. This Technology in Brief will discuss the critical factors of an eDiscovery solution, and will describe Clearwell's approach to fulfilling these factors.

### Overview of eDiscovery Stages

The eDiscovery process is not monolithic but is a complex workflow comprised of many stages, each of which has its own sub-stages. In its simplest form it represents three major tasks: 1) identify custodians and issue litigation holds, 2) collect and preserve data, and 3) analyze, review and produce data – while maintaining chain of custody

throughout the entire process.

Let's break down the individual stages that comprise these tasks. These include Preparation and Identification, Collection and Preservation, Analysis and Processing, and Review and Production. Note that these stages impact more than just the Legal department and outside law firms. They also impact IT and Compliance, which is exactly

why we argue for interdisciplinary teams.

### **Preparation and Identification**

The *preparation* phase starts with enterprise records management. Strictly legal approaches to eDiscovery do not always grant the importance of records management in eDiscovery, but good records retention is vital to cost-effective and accurate eDiscovery actions. Successful eDiscovery depends on tracking and validating data creation, modification, and disposition within potentially relevant data sets.

*Identification* is the process of locating and recording all data locations that might impact an eDiscovery action. In the near-term, identification technology performs this action on-demand. In the long-term, best practice is to do this proactively using classification software. These applications crawl through multiple storage destinations to create an initial baseline index. They then run transparently in the background to keep indexes current.

### **Collection and Preservation**

The *collection* stage enables reviewers to search for potentially relevant data while protecting file type, content and chain of custody. Collection should include not only content matching search terms but also metadata, which is increasingly requested at the meet-and-confer. The collection phase should also protect native format for later production.

*Preservation* starts early in the eDiscovery

process when the corporation has reasonable expectation of a lawsuit. Ideally the corporation should have a defensible and consistent data deletion policy in place, which it only suspends on potentially relevant data once the expectation clock starts ticking. In actual practice, corporations frequently “keep everything” out of fear of deleting the wrong thing. This makes it extraordinarily difficult to order meaningful preservation, and is a clarion call for well-managed records retention in the enterprise.

### **Processing and Analysis**

*Processing* collected data used to be a simple matter of creating load files for export to a law firm or service provider. Given the explosion of data volumes and attendant high costs, corporations have begun to perform this stage in-house. However, the demand for culled but defensible results means that the processing stage has become much more sophisticated.

The *analysis* process helps reviewers to determine the relevance of collected data early in the eDiscovery process. Analysis software commonly provides visual screens where viewers can quickly understand how many times given search terms appear, in what files they appear, and how meaningful the terms are in relation to other collected files. Analysis gives viewers a strong summary of the case during the early case assessment process and before the met-and-confer.

### **Review and Production**

*Review* is the most manually intensive of the eDiscovery stages. Reviewers are charged with viewing responsive documents for relevance and preparing them for production while protecting sensitive information. On top of all this they must justify and verify review decisions. The better the processing and analysis work has been done, the more manageable the review stage will be. But it still remains a staggering responsibility.

The *production* stage can be challenging as well. Even a small company can produce thousands of potentially relevant emails and documents. And in a larger case the number of pages to be produced scales into multi-millions of documents. On top of the sheer time and expense of producing these many pages, parties must maintain security for sensitive data.

### **A High Price Tag**

On top of the demands and complexity of the individual stages, moving data between them is time-consuming and costly. The response of the eDiscovery vendor community has been to integrate their products and partner where they can, but true end-to-end eDiscovery does not exist today. The result is a complex, multi-product workflow that requires frequent data movement and adds to the risk of spoliation and chain of custody issues.

This is a particularly difficult challenge in light of the most expensive and time-consuming eDiscovery stages: Processing,

Analysis and Review (PAR) followed by Production. The lack of efficient and integrated eDiscovery workflow products results in a hefty price tag for these critical processes.

In a typical case, attorneys direct IT to collect emails and other files from 8 custodians. IT collects 400GB worth and copies them to Legal. The project takes staff time from Legal and IT and also costs storage and networking resources, for a grand collection total of \$10K.

Legal takes the collection and subjects it to culling, processing and relevance searches. Culling prior to processing might reduce the 400GB to 90GB. But the processing cost alone for 90GB is approximately \$90k, given a market processing rate of about \$1000 per GB. De-duplication and the application of search terms can further reduce the data to 25GB.

Now comes review time – the single most costly stage of the entire process. Reviewing the 25GB (or 200,000 documents assuming 8,000 documents per GB) costs the company \$150k for reviewer time, figuring a conservative average cost of \$75/hr for attorney review. This gets more expensive the higher up the rank of the review team. Finally 4GB is actually produced at a cost of \$4k. Now rinse and repeat -- and you're talking about the bulk of the expenditures for the eDiscovery workflow process.

## Introducing the 5 Critical Factors

In order to beat these high costs, companies must be able to integrate post-collection stages into a comprehensive and cost-effective eDiscovery workflow. By choosing a solution that offers this level of efficiency and cost-saving, the enterprise can slash that \$200K+ price tag by 80% or more.

This level of dramatic savings requires an eDiscovery platform that contains 5 critical factors: 1) integrated eDiscovery platform, 2) comprehensive processing, 3) strong search and analysis, 4) efficient review, and 5) flexible production.

### Critical Factor #1: Integrated eDiscovery Platform

There may not be an end-to-end solution for the full eDiscovery process, but that does not mean the enterprise should be using narrow point products for every eDiscovery stage. With multiple products, complex eDiscovery operations and data must be moved across disparate software tools. This increases time and cost, wastes network resources, and risks spoliation. Retaining related tasks on an integrated platform lowers much of the cost and associated risk.

It is particularly important to use an integrated solution for processing, analysis, review and production – the most costly stages of the workflow. By using an integrated workflow, the enterprise leverages work done at previous stages and keeps time-consuming data movement to a minimum. And operating the same platform from

processing through production cuts data export and ingestion time considerably, and preserves chain of custody.

### Critical Factor #2: Comprehensive Processing

Processing whittles down the initial data collection to create a single, searchable set of documents for analysis and review. Key processing requirements include analytics, filtering, and comprehensive file support:

1. *Analytics.* Analytics is the ability to holistically view the collected ESI with visualization tools that display a variety of views. Common ways to analyze the data include by custodian, file type and timeline. This ability enables the enterprise to know that they have collected all the potentially relevant ESI which lowers the risk of sanctions. This process also analyzes the total volume of these files. This helps reviewers to better plot the time and cost of downstream eDiscovery stages.
2. *Filtering.* Filtering removes irrelevant ESI prior to the expensive processing stage. Relevance factors include metadata properties such as custodian, date range, file type and file size. This type of filtering commonly removes a good 30% of collected data, which cuts down significantly on processing and review costs.
3. *Comprehensive file support.* Amidst all the enthusiasm for smaller review sets, it

is vital that the review set be complete and accurate. The integrated platform must support a very large set of data types including not only native applications and formats, and also container files and foreign language documents. Attorneys must also be able to justify the completeness of the review set, so the platform must account for each individual document that it processes, including those it excludes. An accurate and defensible process is vital as the enterprise moves to take over eDiscovery tasks from service providers.

### **Critical Factor #3: Strong Search and Analysis**

Search and analysis powers early case assessment (ECA), the process of quickly analyzing collected ESI early in the eDiscovery process. The earlier and more accurate the better, allowing attorneys to build an effective game plan early on for settlement, countersuit or trial. Requirements include:

1. *Defensible search.* Given recent court cases such as *Victor Stanley v. Creative Pipe*, searchers must defend their searches with testing, sampling, iterative refinement, and recorded search queries, as recommended by the Sedona Conference. Without technology-aided defensible search, this is manually intensive and rather impossible in the face of large ESI collections. Poorly tracked and non-defensible search procedures attract court sanctions, and

risk losing attorney-client privilege and work product protection.

2. *Complex query support and analytics.* In stark contrast to very simple search queries used in enterprise search products, eDiscovery search products must support highly complex search strings, be able to run multiple queries simultaneously, and have the ability to quickly analyze results. Analysis is critical for ensuring that search terms are not under-inclusive or over-inclusive.

3. *Foreign language identification.* English may be the *lingua franca* of the business world, but it will never be the sole language for discoverable documents. Multi-language capability should use linguistic and heuristic features to identify foreign language documents so they can be analyzed by experts. This offers strong benefits for today's multi-national business environments.

### **Critical Factor #4: Efficient Review**

Review is the single most time-intensive and costly stage of the entire eDiscovery process. This is why document search and analysis as well as the review product itself can dramatically affect the cost and efficiency of this demanding stage. Critical review requirements include native viewers, custom tagging and redaction. Requirements include:

1. *Native viewer.* This enables reviewers to

perform detailed document review without first converting to TIFF and losing the ability to search documents and view hit highlights during review. This also saves review time by not having to install and launch native applications on reviewer workstations.

2. *Flexible tagging.* This process codes or “tags” one or more documents during review and details how each document is to be treated for production purposes. The review product should have the ability to customize tags, sub-tags and reason codes. The tagging workflow should also help reviewers to reduce errors and optimize the user interface, cutting down on the number of mouse clicks reviewers must make.
3. *Redaction.* Redaction features allow reviewers to protect sensitive information from inappropriate viewing. This is where redacting using a Native Viewer can accelerate the process. In this scenario, the reviewer inputs privileged terms in a search which returns highlighted terms throughout searched documents. The reviewer assigns redaction to the highlighted terms in one fell swoop. Other features include providing custom reason codes for defending redaction decisions. The platform should also be able to auto-redact common sensitive information such as social security or credit card numbers.

### **Critical Factor #5: Flexible Production**

Production may seem like an afterthought to review, but at thousands of pages per case it can be a huge time sink of its own. Flexible and powerful production enables corporations to significantly lower production-based time and risk. Requirements include:

1. *Image-based production.* TIFF remains the format of choice for many productions. Production products should also retain redaction marks on these imaged files, and tagging where appropriate.
2. *Bates stamping.* Production requires the ability to apply Bates stamps and headers, footers, and watermarks on imaged documents. Best practices include automating the sequential stamping or labeling of produced documents.
3. *Rolling productions.* eDiscovery is rarely confined to neat and finite stages. Even as review starts on one set of ESI collection and analysis may be ongoing on more. The platform should support on-demand productions where parties produce multiple sets of ESI throughout the eDiscovery process.

### **Clearwell and the Critical Factors**

Clearwell already has an excellent reputation and widespread deployment within corporations and law firms in the eDiscovery market. With the 5.0 version of its eDiscovery platform, the company adds full-

blown review and production modules and expands its processing and analytics capabilities. Clearwell's integrated offering streamlines the most expensive and time-consuming of the eDiscovery stages, and lets users consolidate the post-collection eDiscovery process on a single platform.

### Clearwell 5.0 Features

Clearwell's goal is two-fold: minimize the movement of case data across multiple tools while improving the efficiency and defensibility of the eDiscovery process. The new release has a raft of new and expanded features to accomplish the mission. Here are a few of the highlights from Clearwell's Pre-Processing, Review, and Production modules.

- **Pre-Processing Module:** The module uses interactive analytics and filtering to a) audit the set of collected data and b) intelligently shrink the size of data sets to be processed. It interactively filters by custom and known file lists as well as various metadata. Analytics visually summarize document set characteristics and perform detailed analysis by custodian, timeline, and file type. Analytics enable IT and legal teams to scope the eDiscovery effort and better prepare for the Meet & Confer.
- **Review Module:** Clearwell does both manual and auto-redaction, uses reason codes, and preserves the redaction work in production. Tagging is another critical capability in review platforms. Clearwell offers Decision-Tree Tagging by

individual documents or document sets (e.g., email discussion threads). Decision-Tree Tagging channels the review process through key decision points, reducing errors and speeding review. Additional capabilities include native viewers for hundreds of applications along with hit highlighting and in-document search support. Clearwell also enables nested folders to manage multiple review sets and has redesigned its user interface to support highly rapid review.

- **Production Module:** Clearwell can produce documents in PDF or TIFF on demand and preserves redaction marks on the imaged files. Production options include must-have features such as Bates stamping, watermarking, and custom headers and footers. The module supports rolling productions, tracks the progress of each production for real-time viewing, and lets users view document sets as they appear in their exported format.

### Taneja Group Opinion

We are big believers in the value of the early eDiscovery stages -- information management, collection and preservation have real implications for compliance and vastly improved data retention policies. However, these are longer-term projects that require time to both get buy-in from different constituents and to implement. In contrast, post-collection eDiscovery represents the most expensive processes and the highest

attendant cost savings. It is also an area where real, hard-dollar ROI can be realized almost immediately. This is why we encourage the enterprise to focus on integrated platforms that provide a consistent workflow and a rapid, accurate, and defensible search and review process. Businesses can realize huge cost savings from highly efficient and cost-effective eDiscovery procedures that integrates pre-processing through production.

Choosing an eDiscovery platform is not a simple decision. Since the highest eDiscovery costs come from the review process and its adjacent stages, customers are clamoring to reduce related costs. Of course eDiscovery vendors are only too happy to oblige, making for a crowded analytics and review marketplace. Buyers must carefully evaluate these offerings. Winning products should

offer sophisticated and integrated pre-processing, review and production; be easy to use for a wide variety of users with varying skill levels; and should be rapidly deployed. Buyers should also hold out for real corporate references, not trial deployments.

Clearwell already provides a widely deployed processing and analytics platform and could have rested on its laurels. Fortunately it did not, but instead chose to build its existing competency with bold new review and production features. The result is a highly integrated solution: a single powerful platform for the critical post-collection eDiscovery workflow. Corporations needing to control review-related eDiscovery costs should take a long look at Clearwell's expanded capabilities.

---

***NOTICE:** The information and product recommendations made by the TANEJA GROUP are based upon public information and sources and may also include personal opinions both of the TANEJA GROUP and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. The TANEJA GROUP, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.*