



Consolidating EDD Workflow with Next-Generation Software

New Software Solutions Streamline Processes Inside & Outside the Corporate Firewall

By Nicholas Croce



Introduction

As the surge in electronic information intensifies, the legal community is seeking new tools to manage the exploding costs and risks of electronic data discovery (EDD) within the mandates of the updated Federal Rules of Civil Procedure (FRCP). A key imperative is to address the need to narrow EDD datasets and apply efficiency to the EDD workflow, from beginning to end, so that costs can become more predictable. The EDRM model developed by Socha-Gelbmann¹ outlines nine separate EDD phases, from information management to presentation, of which two areas represent the highest costs: 1) processing and culling raw data, and 2) document review, especially when taking into account attorney review labor costs.

Software applications are now being introduced to address both of these areas by 1) more efficiently managing collection and pre-processing data, and 2) streamlining and prioritizing document review. Additionally, these tools are focused on integrating multiple EDRM functions in two core areas: operations inside the corporate firewall (information management, preservation, and processing) and functions supporting outside law firms (analysis, review and production).

Addressing Cost Containment & Risk in a Changing Environment

With the pervasive deployment of desktop computing and email, the amount of information generated by corporations and individuals continues to accelerate. With 100 billion emails² generated daily, and the average desktop storing 40 million pages of information and digital rich media (voice mail, video), corporations must now face the challenge to manage and control the information flow within their enterprises. The sheer volume of information combined with legal preservation requirements to avoid spoliation and maintain chain of custody have driven up the costs of the discovery process and increased the risks of finding and reviewing all relevant discovery information.

An additional result of the proliferation of electronic data is the explosion of fees associated with attorney review. Attorney fees associated with the legal review of electronic documents are projected to be 6 or 7³ times the level of electronic discovery spending. Forrester Research⁴ projects that eDiscovery technology spending will surpass \$4.8 billion by 2011; however, based on eDiscovery spending projections of attorney fees for litigation review—approximately \$10 billion in 2006—and the combination of both review and technology fees going forward, eDiscovery spending will exceed \$33.6 billion by 2011. Corporate management will not want to support this expanding level of spending on non-revenue generating costs—it is clear that they will seek containment measures, including new technology-based solutions to manage and reduce these potentially runaway costs.

¹ Launched by George Socha and Tom Gelbmann in May 2005, The Electronic Discovery Reference Model (EDRM) Project was created to address the lack of standards and guidelines in the eDiscovery market. For further analysis of the EDRM project, visit www.edrm.net. For the complete Electronic Discovery Reference Model, please see page 3.

² Peter Lyman & Hal R. Varian, *How Much Information?* (2003) <http://www.sims.berkeley.edu/how-much-info-2003>.

³ Multiples provided by Michael Clark of EDDix Consulting.

⁴ Forrester Research Report: *Believe It—eDiscovery Technology Spending To Top \$4.8 Billion By 2011* by Barry Murphy—December 11, 2006

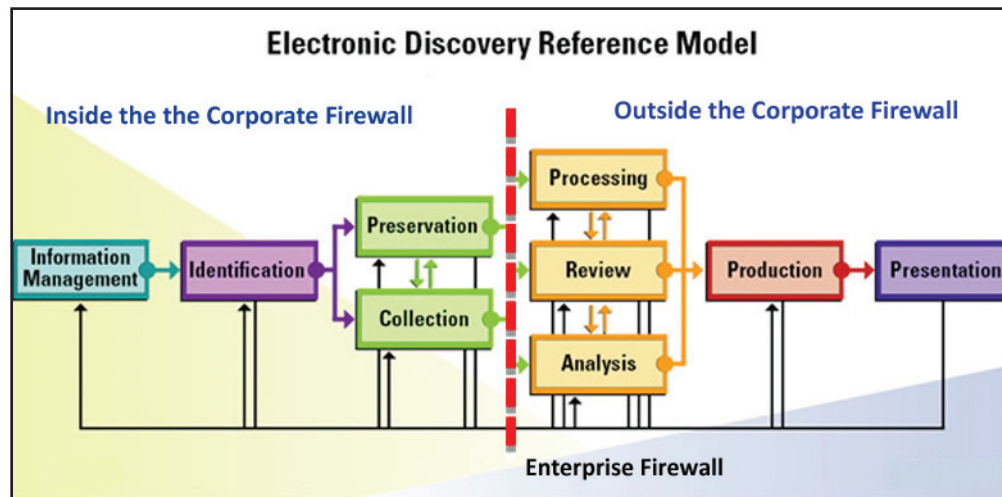
Consolidating EDD Workflow with Next-Generation Software

The corporate legal team is responding to this challenge in conjunction with their law firm partners. Their shared goals are:

- *Minimizing the amount of information to be collected and processed.* The biggest direct cost in eDiscovery is processing vast amounts of collected data. This cost now exceeds \$1 billion—on average the cost is \$1,000 to \$1,500 per GB. Systems and protocols that minimize the amount of data to be collected and processed represent significant potential cost savings. In addition, by using new forensics collection and preservation tools that pre-process data and employ early analytics in the collection stage, general counsel can better minimize the data produced for review and defend the “reasonableness” definitions outlined in the new FRCP.
- *Making the review process more efficient and cost-effective.* Overall, the largest addressable costs in litigation are the fees for legal professionals executing the review process. Forrester Research⁵ has stated that “because lawyers are expensive knowledge workers, review applications that let them quickly review evidence and formulate legal strategies will be extremely important.” New conceptual review tools represent the largest potential cost savings in eDiscovery by speeding up the review process and increasing the number of documents reviewed per hour. Reviewers can also more quickly prioritize and cluster information to determine whether or not data is relevant, privileged, or needs to be produced.

The EDRM Model and Current EDD Workflow Practices

The Electronic Discovery Reference Model (EDRM) outlines multiple workflow categories for the management of electronic data discovery and has become an accepted format for the discussion and implementation of EDD management programs. These “process” areas have traditionally been managed through a services-driven mix of consultation and specialists who apply services and software tools to specific areas such as forensic collection, processing and traditional review.



The rapidly changing dynamic of eDiscovery software and services makes it an appropriate time to discuss how the EDRM model may evolve over the upcoming months and years.

⁵ Forrester Research Report: *Believe It—eDiscovery Technology Spending To Top \$4.8 Billion By 2011* by Barry Murphy—December 11, 2006

Consolidating & Automating Processes Inside the Firewall

The EDRM model can effectively be seen as two “sets” of functions: those that are performed inside the corporate enterprise firewall and those that are performed externally once data has been preserved and extracted from the corporate data store.

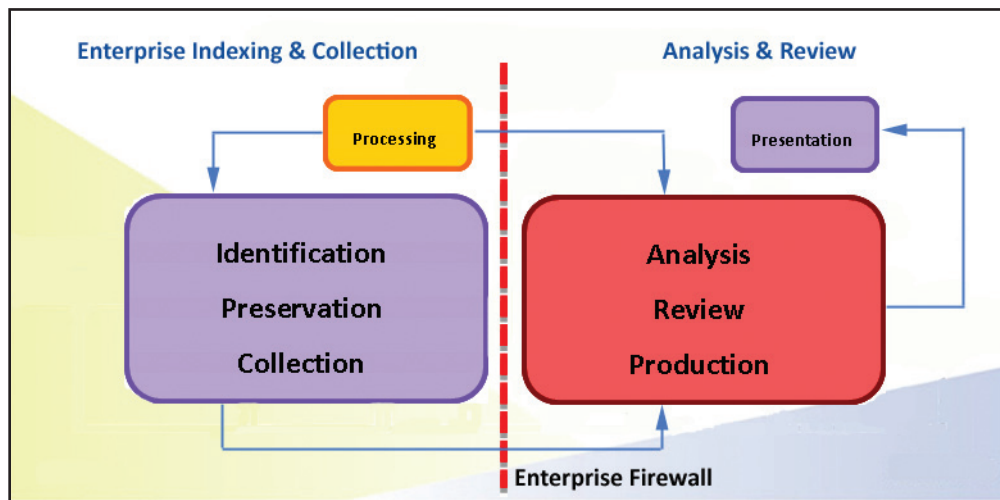
The first set of functions, which include information management, identification, preservation and collection, has required significant assistance from outside experts and consultants to develop case-by-case strategies. Traditionally, the general counsel’s office has managed this process with outside counsel to ensure that evidence was being managed in a legally sound manner. This has often resulted in (and continues to result in) the over-collection of data, which may not provide actionable insight into key evidence. At the same time, corporate data, in both active and backup formats, tends to be dispersed in a variety of formats, creating a situation where significant time and expense is required to collect the data. Once collected, the data then requires “processing,” a set of functions that includes de-duplication, extraction into a reviewable format, and culling via date ranges, custodians and keywords.

In an effort to make collection and processing more efficient and to automate this stage, corporations are deploying new technologies for improved data extraction, as well as pre-filtering during the collection process. A key focus is on access to both active and stored data across the enterprise, efficient collection and search of the same data, and proactive culling and pre-analysis of the data. Corporate IT and Information Management groups are setting up direct linkages to global data and network stores, synchronizing desktops and introducing mobile connections to enable rapid collection. Simultaneously, they are deploying active indexing technology to monitor compliance, perform real-time data warehouse search, conduct internal investigations and risk management analysis. Proactive processing is coming online, where new software tools (via both installed software and hardware appliances) are performing traditional processing to cull data via de-duplication and date range/keyword filtering and metadata extraction during the preservation stage. New software tools can also upload the preprocessed data directly from the corporation, eliminating substantial costs and speeding the process.

Over time, these new integrated tools will help to manage electronic discovery workflow within the enterprise, integrating the four EDRM functions—information management, identification, preservation and collection—into one phase, “Enterprise Indexing and Collection.” This will also result in bringing the bulk of the processing stage within the firewall.

Legacy data sources, backup information or data will still require special handling and may still be executed by external vendors. In the future, these areas will also be integrated internally with the corporation’s system to ensure future productions do not require redundant efforts. By incorporating these methodologies, the scope and resultant costs from EDRM phases inside the firewall will be significantly lowered, with a resultant effect on the reduction in costs of downstream analysis and review.

Consolidating EDD Workflow with Next-Generation Software



Applying Analytics for Faster Review Outside the Firewall

Until recently, review, analysis and production have been executed by outside counsel on a “brute force” basis using litigation support software supporting Boolean keyword searching followed by linear review of one document at a time. Five years ago, with less data to review and first generation tools, this was effective and efficient. However, as datasets continue to expand, the overall inefficiency of the process has called for new tools to more precisely identify key documents and accelerate the review process. With the addition of analytical applications such as concept clustering, email thread analysis and other tools being added to the review toolkit, legal review can now be accelerated by a factor of five or more, saving the corporation significant attorney review fees while also enabling reviewers to find important evidence early in the review process.

Through these new tools, analysis, which has often been a post-review assessment stage, is being transformed into an early-stage activity. Analysis can now be a highly strategic function with the application of conceptual searching to identify key documents much earlier in the process, allowing counsel to conduct an upfront strategic assessment and to negotiate with confidence in FRCP mandated meet & confer conferences.

Law firms, which have been overwhelmed by the amount of data to review, are embracing new analytic review tools so that they can focus their resources on the strategic aspects of the case while also saving significant costs for their corporate clients in the review process.

Conclusion

Driven by the need to limit datasets, manage cost and identify key documents more quickly, new software technologies are beginning to transform and integrate EDD workflow. From a nine-stage process, EDRM model phases can be grouped within two core areas—inside and outside the corporate firewall. By combining upfront control of corporate data within the corporate enterprise with the transformation of analysis into a strategic function, parties involved in large litigation matters will be able to manage risk and cost more effectively. Implementing these new protocols and technologies will enable corporations and their law firm partners to “get to the heart of the matter” more quickly, allowing litigation to be decided on the merits of a case, rather than the cost of electronic discovery.

About the Author

Nicholas Croce, President of Inference Data, has led the creation and development of Inference, the company's next generation analytics software for electronic discovery. Prior to joining Inference, Nick was the president of DOAR Litigation Consulting, a leader in electronic discovery and courtroom technologies.

About Inference Data

Inference Data provides corporations and law firms with next-generation software for analytics-driven assessment, meet-and-confer preparation and accelerated legal review. The Inference product, also called Inference, applies conceptual search and analysis to prioritize datasets and find key documents for faster and more cost-effective review. Inference, a web-based solution, is distinguished by its highly scalable Autonomy architecture and a user-intuitive, configurable workflow. Inference Data is privately held and headquartered in New York, NY. For more information, visit www.inferencedata.com.