



Smarter Discovery™

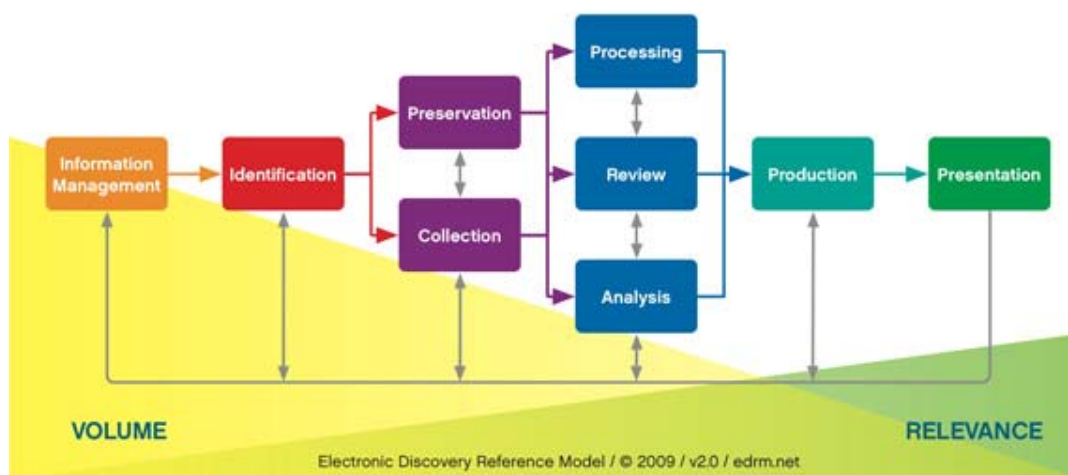
EDRM In A Week

18 TB from Collection through Review in 7 Days

Introduction

The Electronic Discovery Reference Model (EDRM) developed by the EDRM organization is the de facto industry standard approach to eDiscovery projects. The multi-step, multi-level process could seem daunting to even the best eDiscovery project manager. As discovery technology has evolved, seasoned service providers have perfected the execution of the EDRM approach. Expertise and technology has allowed the EDRM model to effectively be executed under tight project deadlines. This paper outlines the process that D4, LLC has developed and used to complete time sensitive eDiscovery projects against the EDRM. The specific case study discussed in this paper provides an overview of a project that required identification through review in seven days.

Electronic Discovery Reference Model



A Phased Approach

A leading law firm had a challenge. The courts mandated the discovery and review of data contained on 78 backup tapes within four weeks. The 78 tapes contained 18 TB of data that needed to be catalogued, indexed and searched. Documents with key word hits were then extracted from the tapes, processed and migrated to the review environment for attorney review. The tapes contained both email archives and native file backup sets. D4 collaborated with their client and the project was broken into two phases:

Phase 1 – Sample Set:

A sample set of the tapes would be catalogued, indexed, searched and deduplicated. The data matching the search criteria would be extracted and migrated to an online review tool. Based on the results of this sample the search terms and data ranges would be refined. This sampling process gives insight into the landscape of the data contained on the tapes and allows the complete processing phase to be streamlined.

Phase 2 – Complete Set:

After the sample set of the tapes was reviewed, the balance of the tapes would be catalogued, indexed, searched, extracted, deduplicated and migrated to review by the legal team.

EDRM Step by Step

Each step of the EDRM process for this project is detailed below along with the time taken and the amount and type of data discovered.

Day 1: Sample Set: Information Management, Identification and Collection

D4 received the set of 9 sample tapes in a box from their client. The original media was checked into D4's evidence tracking system and stored in a secure evidence locker. The chain of custody was maintained from data check in, through all work performed, to final return of media to the client. The sample set of 9 tapes contained 3TB of electronic stored information (ESI). These tapes were indexed, searched, extracted and migrated to review within a 24 hour window using Index Engines automated tape indexing platform.

Using Index Engines, D4 was able to generate a searchable index of the content without first restoring any of the ESI from tape. This process streamlined the identification of relevant data and allowed D4 to present to their client a profile of the tape content as well as detailed knowledge of ESI based on the search parameters. In the past the ESI would have to be restored before processing, adding significant time and expense to the overall project.

Stats: 3TB of ESI (9 sample tapes) indexed and made searchable.

Day 2: Sample Set: Processing and Review

The D4 project team worked with the client to identify and optimize the keyword and date range selection process on the second day. Keyword and date searches were tested and revised over the course of the day. Using Index Engines to query the indexed tape data, D4 was able to identify data on a rolling basis and deliver the processed data to the review environment for client inspection. Once the query terms were finalized, 67GB of the 3TB was determined to be relevant.

Using the Index Engines search capability, all ESI could be queried based on metadata such as date ranges as well as content. Additionally Index Engines also generates MD5 hashes of the content allowing for automatic deduplication of the ESI as it is queried.

Stats: 67GB of ESI was extracted from 3TB of data on tape. (~2% of the sample tape content).

Day 3 – 5: Full Set: Collection and Processing

Based on the initial sample set from the first 9 tapes, key word searches were finalized and the decision was made to index, search, deduplicate, extract and migrate to review the remaining 69 tapes. The 69 tapes contained 15 TB of ESI and indexing occurred over the next three days. Once the tapes were indexed the query was performed. The resulting set of relevant ESI that was then extracted from tape amounted to 157 GB or about 1% of the total tape content.

A critical advantage in collecting ESI from backup tapes was the use of Index Engines technology. This technology allows the direct indexing of the tapes, search of the content and then extraction of relevant ESI. This process avoids full restoration of tape content which has been the traditional method. Since less than 2% of the content was determined to be relevant, Index Engines allowed extraction of this 2% rather than traditional processing that requires restoration of 100% before determining what is relevant. The use of this technology was the determining factor in the project's quick turnaround time.

Stats: 69 tapes containing 15 TB of ESI indexed and searched. 157 GB of ESI (about 1% of total tape content) was extracted from tape for review.

Day 6 – 7: Full Set: Review/Analysis/Production/Presentation

D4 used Index Engines to process all 78 tapes that contained a total of 18 TB of data and ran it against search terms and a date filter within 5 days. D4 extracted the files that had key word hits, processed the files and hosted the data for review. Using the metadata and content search criteria, D4 reduced the data set from 18 TB to 224 GB of potentially responsive data – a reduction of the data set by over 98%. The data was then loaded into a review platform, and uploaded for the client's review. All of this took five days to complete. D4 hosted the culled and deduped data in Relativity for review. D4 worked with their client to develop a prioritized review using Relativity Analytics. The combination of the Relativity review tool powered with Analytics enabled the firm to perform concept searches to better understand the data set and reduce it from 224 GB to 54 GB. Concept searching delivered search results that consisted of conceptually similar documents. These groups of conceptually similar documents were batched to reviewers. Reviewers were able to maximize hourly document throughput and provide more consistent coding within groups of conceptually similar documents.

Categorization, powered by Relativity, also enabled the firm to identify documents that were likely non-responsive based on their content, and prioritize the review. The review team identified a population of 50 non-responsive documents and fed those documents into the categorization engine. The categorization engine analyzes the content of the documents and delivers conceptually similar documents. This process can be iterative. The results can be reviewed and a second set of documents can be submitted to the categorization engine to improve results. This allowed the review team to identify likely non-responsive documents and make them a lower priority review item.

Stats: 224 GB of data was loaded to review and 54 GB was found responsive.

Summary

D4's client, a large law firm, had a four week window for this eDiscovery project. Counsel was able to hit their deadline and stay within budget due to the speed and depth of the services provided by D4 experts utilizing the Index Engines solution. The table below summarizes how long each step of the EDRM model took for this project and how much data was processed.

EDRM Process Steps	Time and Data
Information Management & Identification	1 Day 9 tapes were indexed containing 3 TB of data
Preservation/Collection	1 Day Search terms were decided and querying of 9 tapes with 3 TB of data was completed 67 GB of data matched the search terms
Processing	3 Days 69 tapes with 15 TB data were processed Resulting set of 157GB was responsive to search criteria.
Review/Analysis/Production/Presentation	2 Days: After deduping the sample and second set of data, 54GB was extracted and hosted for further review
Total Project	7 Days 78 backup tapes 18 TB of data contained on the tapes 54 GB was found responsive and held for review

This project is an example of how tape discovery can follow the EDRM approach and still be accomplished in timely, cost-efficient manner. D4 was able to quickly index and identify relevant data and import it into a collaborative review system for rolling client review. Rather than multiple starts and stops between each reference model step, Index Engines allowed D4 to seamlessly transition from one step to the next, for both the sample set and the full population of tapes. The smooth handoff between tape data extracted by Index Engines into Relatively with Analytics allowed an efficient and prioritized review.

About D4, LLC

D4, LLC is a national discovery management and litigation support organization. Founded in 1997, D4 operates two data centers nationwide with consultants and offices from coast to coast. D4 provides professional, defensible, and cost-effective solutions for e-Discovery. D4 solutions include data preservation, collection, reduction & analytics, pre-review, processing, hosting, and production services. Our full range of services and state-of-the-art technology allow corporations, law firms, and government organizations to solve their most time sensitive e-Discovery and document production obligations while ensuring the highest quality, defensibility, efficiency and security standards. For more on D4, please visit www.d4discovery.com.

About Index Engines

The patented Index Engines discovery platform is the only solution on the market to offer a complete view of electronic data assets. Online data is indexed in-stream at wire speed in native enterprise storage protocols, enabling high-speed, efficient indexing of proprietary backup and transfer formats. Index Engines' unique approach to offline records scans backup tapes, indexes the contents and extracts relevant data, eliminating the time-consuming restoration process. Index Engines provides the only comprehensive discovery platform across both online and offline data, saving time and money when managing enterprise information. For more information on Index Engines, please visit www.indexengines.com.