



**AccessData**<sup>®</sup>  
*A Pioneer in Digital Investigations Since 1987*

AccessData Corporation

# Using AccessData Lab to Perform Complete eDiscovery Processing While Mitigating Risk:

A Guide Providing Assurance and Protocol to  
Third-party EDD Providers

White Paper

**Contents**

Introduction ..... 1

The State of Electronically Stored Information (ESI) Processing Today ..... 1

    Non-forensic-based eDiscovery Processing Tools ..... 1

    Forensic-based eDiscovery Processing Tools ..... 2

Addressing the Chain of Custody Element ..... 3

Utilizing Image-based Processing versus Loose File Processing ..... 4

Scaling the Use of eDiscovery Processing Technology for Speed while Mitigating Risk and Liability ..... 5

First Pass Review: Conceptual Breakdown and Value-added Service ..... 6

Conclusion ..... 7

## Introduction

The evolution of eDiscovery processing—or the handling of electronically stored information (ESI) from identification and preservation, all the way through processing and production—has been devoid of standardization. Third-party service providers have historically relied on an assortment of software packages that were singularly based on the premise of capturing metadata and tiff images. The problem with using these "jack-of-all-trades, master-of--none" software packages is that they often ignore the legal implications of how the client data is handled and processed. A simple operator error of copying the client's files to a production server so they can be processed could result in a charge of data spoliation.

Forensically sound and legally defensible procedures, such as maintaining a chain of custody, data control via imaged files, and hashing are often overlooked or never even considered, in order to achieve speed and efficiency for service engagements. For example, working with an imaged version of an original file is almost never considered unless a client has produced one beforehand. This can be attributed to a lack of due diligence on behalf of the vendor who is doing the processing, or the demands of the client who is pushing for a particular method to be used to process the data sets. Either way, the third-party vendor is opening itself to the potential of data spoliation and lost business unless the process is corrected.

Within the last 10 years there have been an abundance of lawsuits against EDD vendors<sup>1</sup>. Yet, at the end of the day, there is still no clear cut processing protocol. Processing tools are constantly upgraded to meet the demands of a challenging electronic environment that is experiencing exponential growth in data creation. The problem now becomes twofold: 1.) How does an eDiscovery service provider manage these large data sets for processing and production, and 2.) how do they minimize their exposure to risk in using certain eDiscovery processing tools?

The answer to this problem is an enterprise-class software solution, built on the fundamental forensic principles of handling data. Not only should it be forensically sound, but it should also process files in a manner that scales in relation to the size of the data universe that it is working with. Even though time is money, EDD vendors should not be forced to risk increased exposure to legal liability in the hopes of processing data quicker. It is possible to achieve enterprise-class speed and efficiency while ensuring compliance for your client.

## The State of Electronically Stored Information (ESI) Processing Today

### Non-forensic-based eDiscovery Processing Tools

Many eDiscovery service providers utilize eDiscovery tools purchased "off the shelf" to meet their immediate processing needs. However, often lost in a vendor's tool evaluation is the demand for process standardization and a solution that facilitates risk mitigation.

Several large and small software companies have produced different eDiscovery processing solutions that are composed of a myriad of different capabilities. Among these various software packages are different types of functionality: collection, preservation, review, processing or producing electronic information. These 5 functions are the major components of the Electronic Discovery Reference Model (EDRM)<sup>2</sup>.

The EDRM was created as a guideline to help standardize the process of electronic discovery. However, while many of these tools attempt to accomplish various steps set out by the EDRM, they fail to address the underlying requirements of establishing a standardized, defensible process and mitigating risk.

---

<sup>1</sup> Anthony Lin, "Sullivan & Cromwell Suit Against Vendor Highlights Problems With E-Discovery", *New York Law Journal*, January 7, 2008 (<http://www.law.com/jsp/article.jsp?id=1199441137204>)

<sup>2</sup> <http://www.edrm.net/>

Just because a software package claims it will do one thing, does not mean it will do it in a manner that will minimize risk for a company using the software. Software companies rarely produce a tool that actually facilitates the establishment of a legally defensible process. Furthermore, vendors using a disparate collection of tools will find it challenging to establish a standardized, defensible process on their own. Yet vendors have been known to overlook potential defensibility issues in exchange for speed on the job.

Spoliation of electronic data is the intentional or negligent destruction or alteration of evidence when there is current litigation or an investigation, or there is reasonable anticipation that either may occur in the near future. There are many missteps providers may take that can result in data spoliation if they are not careful. Generally speaking, a third-party service provider will receive ESI from a client, process it accordance with the specifications set by the client, and either initiate a review process or produce the processed ESI in a format which the client can use— for example as a load file for a review platform or in native file format. However, often missing from this EDD processing scenario is any semblance of a control mechanism which allows the received data to be handled in a forensically sound manner. Much of the EDD processing software on the market today is not built on a forensic framework. These products do not take into account hashing principles, imaging techniques or chains of custody within a standardized process. They simply process data.

From receiving the data, to copying it for processing, every move the vendor makes with the client data is an opportunity for failure from a risk management perspective. By not logging the data into the software formally, the vendor may break the chain of custody. Not initiating hash procedures during preprocessing may break the chain of custody as well. The worst transgression of all is processing the client's live data and not working from an imaged copy. By operating on the client's live data, the eDiscovery vendor runs the high, if not probable risk of altering the core metadata of those files and thus changes them from being the requisite "true and exact copies" to being considered by a court of law as different files altogether. Again, just "processing data" may work from a practical data viewpoint, but it fails tremendously in the eyes of the law, exposing the EDD vendor and its client to a great deal of risk.

By establishing the forensically sound protocol of working with imaged copies, hashing out file values, and establishing and maintaining a chain of custody, EDD providers can virtually eradicate the risk of data spoliation. Remove the opportunities for failure by only working with imaged copies of the forensically preserved data, verifying the files as true and exact copies via hash analysis, and maintaining a chain of custody—and the process virtually ensures itself against risk.

## Forensic-based eDiscovery Processing Tools

On the flip side of the processing software coin there is a realm of software that satisfies the Court's interpretation of how to handle and process electronic information. These products actually embrace and are built on the foundation of forensic principles. The advantages of utilizing forensic-based technologies for processing data are relatively well-known at this point due to a slew of court cases highlighting what can happen to a company guilty of spoliation. In addition, these forensic solutions have utilized within large corporate data environments for years, long before eDiscovery was a hot button issue. This means that these technologies have had more time to mature.

Corporate investigators as well as incident response teams have had to mobilize across large infrastructures of computers and servers in response to litigation readiness protocols and have utilized forensic software to do so. Forensic software providers have scaled their technologies to meet the demands of the corporate environment, providing enterprise-class forensic solutions. This enterprise-class forensic technology has now filtered out to the eDiscovery processing arena in response to the need for processing large collections of corporate data.

AccessData, founded in 1987, has made its name in the forensic marketplace as a leader in forensic and cryptography tools. In fact, a majority of information security, computer forensic and EDD service providers use AccessData's flagship technology, Forensic Toolkit® (FTK®). In order to allow law enforcement, service providers and corporate and government FTK users to expand their forensic analysis and processing capabilities, AccessData designed a solution called AccessData (AD) Lab.

Within the context of eDiscovery, AccessData Lab allows the EDD service provider to turn FTK into an enterprise-class processing tool. AD Lab is currently used to amplify existing forensic analysis resources with several Fortune 500 companies, as well as computer forensic labs at the state and federal levels.

AD Lab incorporates the structural integrity and power of the FTK platform, as well as the cutting-edge Microsoft Silverlight technology for "see through" data accessibility and increased speed, all the while maintaining court-required forensic methodology. AD Lab gives eDiscovery vendors the peace of mind in knowing that the software is protecting their own legal interest by facilitating a standardized, defensible process.

Third-party EDD providers do not have to be trained certified computer examiners to utilize AccessData Lab. The upside to using this solution is that it walks non-forensic users through the each phase of data handling. So experienced forensic examiners and non-forensic users alike can enjoy the benefits of using the software. In addition, because the solution wizard-driven and easy to use, the training required to get up and running effectively with the technology is minimal.

## Addressing the Chain of Custody Element

If the client needs to have electronic data presented at some point during a legal proceeding, process dictates they must first have it admitted. In order to do so, they are going to have to address the chain of custody originating from the evidence to be admitted all the way back to the file as it existed at the time of preservation.

What many people fail to recognize is that through this chain of custody process, the litigation support service provider now interjects itself as link in that chain. In order to satisfy a portion of that chain of custody, the vendor must now keep track of the data as it comes to them from the client. Unless an EDD vendor has proactively taken forensic classes and understands this duty, they may not know what this really entails. At a minimum they must create a list of files for inventory purposes, and then hash out the original client files to act as an identification protocol. Without this protocol in place, anyone within the chain of custody would have a hard time saying that the files received by the vendor are true and exact copies of the original files.

This is where AccessData Lab comes into play. Once a new processing case is set up in the software (identifying background case descriptions, time zones, etc), the operator is given a choice of the different data hashing sets to select before even processing the data:

1. **MD5 Hash:** an MD5 hash value is a 32-character value calculated from the file that is unique unto itself.
2. **SHA-1 Hash:** a SHA-1 hash value is a 40-character value calculated from the file that is unique unto itself.
3. **SHA-256 Hash:** SHA-256 is a longer version of the SHA-1 hash system. SHA-256 extends the value to 256 characters for each file.

Although a user of AD Lab does not have to create a hash of each type, the software provides the option to do so. Some forensic labs across the country require SHA-256 hashes in the course of their work, while others consider an MD5 hash to be sufficient. Regardless of the hash type selection, AD Lab allows the traditional EDD processing vendor the same kind of forensic methodology as used by federal- and state-level forensic labs.

It is understood that EDD processing vendors may never aspire to operate their own forensic lab. However, the software affords them the same level of data protection as utilized by computer forensic professionals. AccessData Lab is providing that extra level of protection to the processing vendor by making the hash value computation by default with every operation.

Consider this: Each version of hashing creates a fingerprint of that individual file that enables an EDD provider to say, "The client gave me this file, and before copying or processing it, I have identified it, and it remains in my possession unaltered". The chain of custody is maintained BEFORE processing has even begun and thus satisfies an element of a forensically sound and legally defensible processing operation.

If a third-party provider ever had to go to court and attest that they received and processed a file in accordance with the wishes of the client, they can provide a report from AD Lab that shows exactly that. The importance of providing a hash file before the onset of data processing cannot be overstated, as it affords a level of protection that covers the third-party processor in the event that it is needed.

## Utilizing Image-based Processing versus Loose File Processing

Traditional non-forensic EDD processing software more often than not processes loose files contained within a folder structure or from a structured data repository. Sometimes the original data remains unaltered prior to processing, other times important metadata on the original file is changed. The eDiscovery processing vendor won't know this to be the case until a client remarks back that the trial judge is raising an inquiry into whether a file produced as evidence is a "true and exact" copy. At that point, it is too late for the processing vendor to claim ignorance or to save their good name.

To avoid that situation altogether, AD Lab offers a feature that stays within the forensic framework of handling data by allowing the operator of the Lab software to "Image on the Fly". The user is prompted to image a collection on the fly before processing when they attempt to add evidence to be processed to a particular case. What this means is that the software will read where the files are located, and then create a container file to hold all the files to be processed. This imaged container will act as a safe data repository to which the original files are safely copied, in order to be "processed". This is extremely useful and important due to the fact that it allows EDD vendors the ability to work from imaged media and not the original files. For example, think of the imaged file containing a client's DVD files as a lock box of sorts. This lock box provides another layer of credibility to the methods used by the vendor as the original data can never be associated with data spoliation.

If you were to simply process a client's files without imaging them, you run the risk of an operator either copying over the files to a network location to be processed (thus altering the data, making it no longer a true copy of the original file) or opening the original files on his or her own to be processed. Either way creates a situation where the processing vendor is contributing to data spoliation and can thus be at risk for a law suit. When an EDD provider is working with an imaged copy of the file, they will never touch the original file again and will be working with an exact duplicate of said file. Just one more step of adherence to a forensically sound protocol.

## Scaling the Use of eDiscovery Processing Technology for Speed while Mitigating Risk and Liability

As lawyers have had to deal with the increased burden of reviewing thousands of documents electronically, software developers have had to meet the same challenge as well. Many developers of such tools have been in business for years, but have not had the ability (or choose not to alter their code due to internal restructuring costs) to scale their software upwards to meet the challenges of processing extreme amounts of electronic data.

And it is only going to get worse as more and more data points of discoverable information are found on:

- A) Networks
- B) Desktops/Laptops/Workstations
- C) Cell Phones/PDAs
- D) Tape Backup Systems
- E) SAN/NAS systems
- F) Legacy Systems
- G) VOIP Systems
- H) Structured Data Repositories

By and large, many EDD tools on the market today are designed to do very specific jobs. The problem with many of these tools is that they were built with unintentional limitations. Three years ago, many EDD processing vendors were dealing with corporations imaging entire network servers in response to litigation hold orders. This was due in part to legal teams, motivated by court precedent, being overly cautious about complying with preservation orders. This put a burden on many software producers who now had to handle amounts of data never seen before.

Enterprise-class software is designed just for that situation—built for large-scale data analytics. This level of software is built upon the foundation of accessing, identifying, collecting and processing large amounts of data from a local or remote point of operation. By its very nature, it has to be robust enough to keep up with the demands of the corporate environment. This means it is capable of accessing, copying and analyzing thousands of points of interest on the network map.

AccessData Lab is based on this enterprise computing principle, incorporating the ability to scale proportionally with each processing project. Users of the solution can add as much hardware as they want to their processing environment in order to meet the demands of the data to be processed. AD Lab gives the user the ability to add several "workers" to the infrastructure so as to enable a distributed processing environment. Instead of one server working on a project, the software allows the work to be distributed among existing resources within the vendor's internal infrastructure to decrease the time it takes to fully process a case for a client. It is this new enterprise-class processing capability that allows vendors to offer a value-add to their own clients in the amount of time it takes to handle and process electronic data.

Distributed processing is the only type of processing environment that makes sense in light of how much data exists and how it will keep growing. Even directed and culled down data sets make for large processing jobs. Scalability used within a forensically sound processing operation has been discussed for some time, but it is now a reality.

Mitigating risk while increasing speed means that the software cannot miss a beat when scaling in size to the data. Because AccessData Lab is based on the distributed model of processing, and because large data sets can be evenly distributed against several workers on the network, each file can be given the CPU and memory intensive attention that it deserves. Other tools may try to do that, but often times they are either limited by a single box operation (non-distributed) or the distributed elements of the processing architecture are untried or limited themselves. Either way, the vendor is not getting a true enterprise-class solution capable of dealing with large

amounts of data in an effective manner. The last place a vendor wants to be is in a courtroom explaining how their software platform failed to process a particular piece of evidence because their software "missed" it.

## First Pass Review: Conceptual Breakdown and Value-added Service

With the increase in litigation data volume, especially when considering large data repositories, such as Microsoft's SharePoint and Exchange Server, vendors are increasingly aware that you cannot process everything without incurring a time or monetary cost.

The question becomes one of how to reduce the volume of data to be processed, and ultimately billed out? First pass review within a forensic framework is the answer.

Document retention policies and litigation holds may dictate what is made available for eDiscovery processing, but the software should possess what is called "first-pass review" functionality. This first-pass review should enable the user to do a quick cull down based on certain criteria that reduces a data set from a large amount to a smaller amount as quickly as possible. The resulting smaller response set now becomes the focus piece for the next set of processing.

Conceptually, it is now a value-added proposition to the EDD vendor's clients, simply because it reduces the cost of final processing. However, typical eDiscovery processing software has limitations as to the preprocessing operations that are required to do a first-pass review. Existing software will take in a set amount of data, process the data according to certain specifications and then present the data to be then culled down into a smaller set for other review tools. It is a lot of work, and training, for the vendor to cull a large data set down to a smaller one for the client. System resources are tied up during this process that could be used on other projects. The concept starts to fail as user error is introduced to the process and now the vendor finds themselves in a cost detriment as opposed to a value add. The word 'rework' has become a synonym for 'loss of profits'.

AccessData Lab, besides being built on a distributed processing model that allows for as much or as little hardware usage in the vendor's infrastructure system, gives many options to the user up front that allow for a more complete first-pass review:

- 1) Flagging bad extensions
- 2) Expanding compound files
- 3) Flagging duplicate files and deduplicating based on granular relationship levels
- 4) Include only certain specific file types, or all file types
- 5) Only include files based on a certain file date range or size range

Once those certain criteria are specified, AD Lab will now only include those files into the processing engine for full index and metadata processing. Once those files are processed, the user can now take advantage of a next generation user interface. Utilizing Silverlight, Microsoft's new graphical user face (GUI), EDD processing operators can now quickly drill down and search against this set of data and produce an even smaller set for further review or processing.

This first-pass review is critical in giving the client the most accurate and relevant data set possible to review, as opposed to inundating the client with a large volume of irrelevant files to sift through. Working in tandem with the client, the provider can now take a set of keywords or other search criteria and run fully qualified keyword searches to obtain a quick and efficient first-pass review set of electronic files for further dissemination. AccessData has taken the concept of first-pass review, and transplanted it into an Enterprise-class solution that has been engineered to deal with real-world data sets.

## Conclusion

eDiscovery vendors now have two paths to take in their business model: conduct their business in a risk mitigating fashion by utilizing an enterprise-class, forensically sound solution, such as AD Lab, or continue on a path of utilizing technologies that expose their companies to risk. eDiscovery processing software has evolved slowly in the last 10 years, but companies, such as AccessData are changing the way people do business, due the performance of solutions, such as AD Lab, as well as the risk awareness that the software was built on.

Taking advantage of a distributed processing software model, coupled with extensive reporting and data control allows today's EDD service provider to operate dramatically faster and more efficiently within the sphere of a forensically sound practice. AccessData Lab will allow service providers to handle more cases and larger data sets, while providing value-added services to their clients, such as more targeted culling and first pass review.