

The Art of eDiscovery

Not All Providers are Created Equal

eDiscovery in the Information Age

It's no secret that the expanding volume of electronically stored information (ESI) involved in a typical litigation matter presents a challenge for companies and law firms. In 2000, a typical case might have involved five to ten gigabytes of data. Today, it's common for cases to involve multiple terabytes. Corporations and their law firms must manage this data explosion in a cost-effective manner while meeting court deadlines and discovery obligations.

It is not only the cost of ediscovery that is a challenge—the typical process can place months between data collection and the attorneys' ability to view the data and make critical case decisions. Documents must be collected from disparate sources, processed, and loaded into a review tool—all of which can take weeks or even months—before attorneys can even begin to view the documents. In addition, each data set is different, and often unpredictable. The data may involve hundreds of different file types, all of which require different amounts and types of manipulation and processing before they can be reviewed in a meaningful way.

Even after review finally begins, attorneys may have to wade through a massive volume of information, much of it irrelevant, before they can find the key information that enables smart case assessment and strategy decisions. Paradoxically, because ediscovery is often so unwieldy and expensive, it's more important than ever to be able to assess the risk, merits, and potential costs of a case early in the game.

Now, improvements at the processing stage of ediscovery, made possible by new technologies and augmentations to those technologies, can help solve these problems.

1. Processing is simply faster.
2. Augmenting third party technologies allows processing a wider range of file types automatically, further reducing delays, and allowing ediscovery practitioners to handle even the most complex data with ease.
3. Better search capabilities exist, allowing real-time testing and fine-tuning of keyword searches and other data reduction strategies.

These improvements translate to faster, more economical processing; more effective and efficient data reduction; easier document collection; earlier case analysis; and better preparation for Rule 26 conferences.

Processing in a Traditional eDiscovery Workflow Suffers From Limitations

Before ESI reaches the review stage, it must be identified, preserved, collected, and processed. In the early days of ediscovery, processing was straightforward: it entailed printing out electronically-stored documents and scanning them to create tiff images. Now, processing is much more complicated and time-consuming: data is indexed; meta-data is captured as it existed prior



to processing; data may need to be restored from back-up tapes; individual files and emails may need to be extracted from container files; and certain types of files, such as legacy mail formats, may need to be converted to a different type of file before proceeding. In addition, different processing tools have different file type limitations (i.e. they cannot process all files), different combinations of customizable settings, and differing abilities to handle complex workflows. A data set may involve dozens of different file types, several different time zones, and two or three different operating systems. Further, various strategies can be applied to reduce the volume of data, such as de-duplication and filtering by keyword, concept, date, and/or file type. Such filtering is intended to eliminate irrelevant documents and reduce the volume of ESI that must be reviewed by attorneys.

PROCESSING CAN TAKE WEEKS

This simple fact means that it may be months after litigation arises before attorneys can begin reviewing documents and locating meaningful information about the case. This delay in document assessment is especially problematic because discovery itself has become so costly that avoiding it through settlement can be a good strategy. It is difficult to determine the best settlement strategy before seeing the documents to assess the risk and merits of a case. The time involved in processing can also make it difficult to comply with court deadlines, and limit the ability to act quickly when those deadlines are approaching.

PROCESSING INVOLVES UNCERTAINTY

At the processing stage, narrowing and reducing the data is often necessary so that it can be reviewed timely and cost-effectively. In a traditional ediscovery workflow, that process may involve guesswork. Often, attorneys create a list of keywords expected to be present in relevant documents, and use that list to narrow the data. Problems often arise with this approach. For instance, a particular keyword might have unanticipated results, picking up a large number of irrelevant documents (i.e., a search for "SEC" might

pick up documents discussing college football (the Southeastern Conference)). Or, a particular keyword might have a common synonym or misspelling that often appears in relevant documents instead of the selected keyword, resulting in a large number of relevant documents not being picked up by the search. (ie, A search for "oil spill" will miss documents referring to an "oil leak.")

Thus, the processing stage of a traditional ediscovery workflow can take a long time and may result in a data set for review that is needlessly large with many irrelevant documents, excludes a substantial number of relevant documents, or both. Such issues may not be revealed, if at all, until after attorneys have begun reviewing the documents. By then, it is often too late to remedy such errors.

Providers typically take two approaches to ediscovery. The first is using an end-to-end solution tool that is intended to perform all tasks from processing to review within the tool itself. The second approach is the use of several different tools in a multi-point solution environment. The perks of using an end-to-end solution is that all processes and workflows are housed within one tool making it easy to perform training and move data through each stage of the EDRM. However, the general functionality and workflow of an end-to-end tool will usually highlight one specific area where it shines, while the rest of the workflow actions only produce mediocre output due to tool limitations. This is where the use of a multi-point solution becomes beneficial. It enables the provider to use each tool for its strengths and customization options and then move to the next phase of the workflow, providing superior work product more efficiently. The problem is, that in using a multi-point solution, data frequently must move between several different ediscovery tools before being loaded for review or ultimately being produced. This opens the door to the opportunity for error each time data must be exported and imported into a tool, and without training, can become a very time intensive task.



Lighthouse Middleware: Improved Data Reduction and Analysis Workflows

Lighthouse has been providing ediscovery solutions to Am Law 200 and Fortune 500 companies for more than 16 years. Its seasoned team of legal professionals is assembled from Am Law 100 firms, top-tier ediscovery firms, and corporate legal departments and has significant ediscovery experience.

With the benefit of this experience, Lighthouse has carefully selected the best available technologies and created a proprietary workflow. The result is Lighthouse's Data Reduction and Analysis workflow, which maximizes the benefits that can be gained from improved data processing.

SPEED OF PROCESSING

By using a series of industry leading processing tools, *Lighthouse has processed an average of over a terabyte of ESI per day, completing in hours what would have taken weeks in a traditional ediscovery process.* In other tools, this initial processing phase would have taken weeks. Initial data intake speeds, however, only paint one part of the processing picture. Virtually every processing tool on the market, whether part of an end-to-end solution or a point solution, has limitations on the types of files that can be processed automatically. These files are typically flagged as exceptions and require a technician to manually open and process each of them before a volume of data is complete and can be culled—a costly proposition in terms of time and money. Lighthouse has developed custom automated tools integrated in its workflow that can automatically process these files that would otherwise require manual processing. (ie, A task that once might have taken 16 hours of manual processing now only requires 40 minutes of computer time.)

REDUCING DATA BY 85 PERCENT OR MORE

On average, Lighthouse is able to reduce the data by 85 percent before review, as opposed to 60 percent in

a traditional ediscovery workflow. This can dramatically decrease the cost of review, which is the most expensive phase of ediscovery.

In a case involving 500 gigabytes of data, an 85 percent reduction in data can translate to more than one million dollars of savings over a traditional workflow, even if the rate of review does not increase due to other improvements.

TABLE 1: BREAKDOWN OF DATA REDUCTION AMOUNT, SPEED, AND ASSOCIATED REVIEW COSTS

| Data Reduction % / Remaining GB | # of Documents 7,500 docs/GB | Total Review Hrs. 100 docs/hr | Review Cost \$50/hr |
|------------------------------------|---------------------------------|----------------------------------|------------------------|
| 0% / 500 GB | 3,750,000 | 37,500 | \$1,875,000 |
| 60% / 200 GB | 1,500,000 | 15,000 | \$750,000 |
| 85% / 75 GB | 562,500 | 5,625 | \$281,250 |

EARLIER CASE ANALYSIS

Improvements in data processing enable earlier case analysis, because documents can be processed and made available for review much more quickly. In addition, Lighthouse's data reduction and analysis workflow allows information to be gleaned from the data even before attorney review begins. For instance, Lighthouse can automatically analyze domain and email addresses in the data to generate a report on key custodians or potential deponents that shows, in a simple and graphic manner, who they communicated with and how often.

Lighthouse has found that this can help litigants identify additional custodians that may have been overlooked, understand relationships between key players, interview witnesses, and prepare for depositions. In the following example, Ms. Rangel appears to be a key witness, since she had numerous communications with two key custodians, and a closer look at those communications might be warranted before settlement strategies are devised.

Our early case analysis reporting also provides litigants with a detailed report of all data existing in their matter at any given time. This enables the client to view information such as file types, date ranges, culling reductions and the results of search terms. As noted



FIGURE 1: SAMPLE CUSTODIAN DOMAIN AND EMAIL ADDRESS ANALYSIS REPORT



above, we can also provide hit-by-term reporting that can be used to further refine search terms as well as e-mail domain reporting. By being able to view hit-by-term reporting early in the process, we can easily work with the client to identify and refine terms that may be over or under inclusive in their results.

AUTOMATED, REPEATABLE LOADING OF DATA FOR REVIEW AND PRODUCTION

Lighthouse has developed an automated tool that enables data to flow seamlessly between each of the tools in our multi-point solution workflow. This enables us to reduce the risk of error as well as the amount of time required to move large datasets through to the next stage in the ediscovery processing lifecycle. Using these tools and workflows, we are able to leverage the strengths of a single-point solution workflow with the power of several multi-point solutions without being forced to settle for the weaknesses in choosing only one or the other. The end result is the highest quality work product possible without concerns of time or error to achieve it.

MAXIMIZING THE EFFICIENCY OF REVIEW

Improvements in data processing also allow attorney review to begin sooner. In addition, if the data set is smaller and more relevant, attorney review will proceed more quickly and economically. For instance, with client direction and interaction, Lighthouse can leverage its PrivSmart™ product to help pre-sort and prioritize privileged documents in review. PrivSmart also identifies documents that respond to privilege searches but are likely not privilege, such as footer-only hits and e-mails from counsel between third parties. Additionally, Lighthouse can leverage threading tools to analyze e-mail and suppress lesser fully included e-mails and apply coding decisions made on the reviewed e-mail to all lesser included copies. This reduces the volume of e-mail to review while not excluding it entirely from production. Such pre-sorting means that less work must be done at the review stage and makes actual review time more efficient, economical, and easier to work through.

Improvements in Processing Lead to Benefits at Every Stage of eDiscovery

Improvements at the processing stage of ediscovery, made possible by leveraging best of breed technologies at every stage of the EDRM, can go a long way towards improving these problems. Better processing streamlines culling, searching, and processing, enabling clients to reduce costs and access key information about cases earlier in the lifecycle. This is due to speed, the ability to handle a wide range of file types, powerful search and reporting techniques, and automated and repeatable processes for loading documents to review.

Using leading technologies, and building its own “middleware” augmentations to those technologies, Lighthouse has been able to process data dramatically faster than was possible with traditional ediscovery software tools—on average, nearly 100 times faster. This speed translates into cost savings for companies.



TABLE 2: COMPARISON OF TRADITIONAL PROCESSING SPEEDS VS. LIGHTHOUSE PROCESSING SPEEDS

| Amount of Data | Traditional Process | | Lighthouse's Process | |
|-------------------------|---------------------|-------------|----------------------|--------------|
| | Hours | GB/hr | Hours | GB/hr |
| 40 GB | 33.00 | 1.21 | 0.67 | 59.70 |
| 55 GB | 82.39 | 0.67 | 1.00 | 55.00 |
| 172 GB | 363.55 | 0.47 | 3.30 | 52.12 |
| 2,637 GB (Total) | 478.94 | 0.56 | 4.97 | 53.76 |

In recent projects, Lighthouse achieved the following results, completing in hours what would have taken weeks in a traditional ediscovery process:

This powerful functionality, when combined with a knowledgeable, innovative technical team and an experienced project management team, creates a faster, more economical and more valuable service to corporations and their law firms. These developments lead to five major improvements in ediscovery processing.

1. MORE EFFECTIVE DATA REDUCTION

A central goal of processing is to narrow the data collected down to a focused, relevant set for review. This stage involves filtering the data by keyword and date. New technologies with highly customizable search capabilities and enhanced speed make it easy to fine-tune searches in real time. Searchers can create potential queries by analyzing actual information about the data, such as a report of word count by custodian, and then run different searches and view the results quickly. It may be possible to determine if one search term in particular is resulting in a large number of irrelevant hits, and figure out how to exclude those hits. It is also possible to review a sample of the documents not being retrieved by searches, helping to ensure that relevant documents are not being excluded. Thus, in addition to the speed with which documents are available for review, better data processing makes it easier to have confidence that the final data set for review is small, yet comprehensive and relevant. It also helps ensure that ediscovery

decisions are legally defensible. With new technologies and the right expertise, data processing is completed quickly and the data is reduced by 20 to 30 percent when compared to traditional processing. Data can be reduced even further if the client and service provider create specialized culling rules specific to the company's type of data.

2. EASIER DOCUMENT COLLECTION

Improvements at the processing stage can also make collecting the data simpler.

Because it is so fast and easy to process, sort, and filter the data at the processing stage, less need exists to cull data at the collections stage, which companies may not have the technology, resources or detailed processes to do effectively. Instead, companies simply collect custodians' mailboxes, and irrelevant data can be quickly sorted out at the processing stage quickly and economically.

3. EARLIER CASE ANALYSIS

Because processing and early case analysis takes place in a fraction of the time it would have taken in a traditional ediscovery process, attorneys can start getting into the documents much more quickly. This means that attorneys can assess the risks and merits of a case earlier and make earlier and more informed case and settlement strategy decisions.

4. BETTER PREPARATION FOR RULE 26 CONFERENCES

Federal Rule of Civil Procedure 26(f) requires that parties confer "as soon as practicable" and discuss discovery. Among the topics the parties are directed to discuss is ediscovery: "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced."

In a traditional ediscovery workflow, search terms are often agreed upon at a Rule 26 conference before the terms have been tested. This can have a dramatic impact on the cost of ediscovery as well as on whether litigants are able to get the information they need



from opposing parties. Better data processing allows for testing and sampling of search terms before Rule 26 conferences, so that litigants do not have to go in blind.

The EDRM Project recommends such validation of search strategies and courts are increasingly recognizing its importance. For instance, Magistrate Judge Peck of the Southern District of New York recently admonished lawyers:

[W]here counsel are using keyword searches for retrieval of ESI, they at a minimum must carefully craft the appropriate keywords, with input from the ESI's custodians as to the words and abbreviations they use, and the proposed methodology must be quality control tested to assure accuracy in retrieval and elimination of "false positives." It is time that the Bar—even those lawyers who did not come of age in the computer era—understand this.

William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., No. 07 Civ. 10639, 2009 WL 724954, *3 (S.D.N.Y. Mar. 19, 2009).

5. LESS COSTLY ATTORNEY REVIEW

With improvements in data processing, attorney review can begin much sooner than in a traditional ediscovery

workflow. It also proceeds more quickly, because the data set is smaller and more finely-tuned as a result of the processing. Attorneys will not be hampered by having to wade through a lot of irrelevant data.

Conclusion

Improvements at the processing stage of ediscovery, made possible by new technologies, can go a long way towards improving problems present in a traditional ediscovery workflow. Lighthouse has combined the best new technology with deep expertise to create its data analysis and reduction workflow, which maximizes the benefits that can be gained from improved data processing. Though the benefits are greatest in cases involving large amounts of data, improvements can exist even in small cases. *All of this means more value to the client for fewer dollars spent: in a case involving 500 gigabytes, Lighthouse estimates that its expertise, new technologies and new workflows can save clients well over one million dollars, just to get through the first pass review stage of the case. Better settlement strategies, more efficient prosecution of the litigation, and other downstream benefits may result in even larger savings.*

About Lighthouse

For 25 years, Lighthouse has provided innovative software and services to manage the increasingly complex landscape of enterprise data for compliance and legal teams. Lighthouse leads by developing proprietary technology that integrates with industry-leading third-party software, automating workflows, and creating an easy-to-use, end-to-end platform. Lighthouse also delivers unique proprietary applications and advisory services that are highly valuable for large, complex matters, and a new SaaS platform designed for in-house teams. Whether reacting to incidents like litigation or governmental investigations, or designing programs to proactively minimize the potential for future incidents, Lighthouse partners with multinational industry leaders, top global law firms, and the world's leading software provider as a channel partner.

Connect with us to see how Lighthouse can best support you.

(206) 223-9690 | lighthouseglobal.com | info@lighthouseglobal.com

