“If you think compliance is expensive, try non-compliance”

Former US Deputy Attorney General
- Paul McNulty -
Introduction

In these litigation happy times, eDiscovery is a serious cost center for the modern organization. eDiscovery does not generate direct revenue for the company, disrupts businesses and limits the time and energy that management could otherwise invest in their core activities.

That said, there are few organizations who can afford to do nothing, as the consequences of not being prepared can cost a company considerable more. Multinational organizations deal with legal discovery and disclosure requests for Electronically Stored Information ("ESI") and paper documents on a regular basis. These organizations should look into solutions that can lower the cost of their eDiscovery projects. Gaining control over eDiscovery data is recommended for companies that operate in a highly regulated or highly competitive market that pile vast volumes of multilingual data in multiple storages and that have to handle multiple legal cases per year. For these companies, bringing an eDiscovery solution in house and preparing for potential litigation, drastically reduces storage costs, legal fees paid to outside vendors and lowers IT operations costs.

"Bringing an eDiscovery solution in house, drastically reduces storage costs, legal fees paid to outside vendors and lowers IT operations costs"

Figure 1 – The Electronic Discovery Reference Model (EDRM) showing the various stages of the eDiscovery process

eDiscovery projects are expensive and costs can vary per different pricing models that are offered by various vendors. The costs discussed in this whitepaper are based on publicly available data and experiences of ZyLAB clients.
An overview of publicly available market data can be found in Table 2 in Appendix A that displays the estimated costs of eDiscovery projects for the different stages of the EDRM.

In this white paper we discuss the possible costs saving in the several stages of the eDiscovery process and we will illustrate how you can lower your eDiscovery costs by being pro-active towards organizing and managing information that may be relevant for your cases.

Handling (unstructured) data

Key in reducing cost for eDiscovery, is the way that companies handle their (unstructured) data. Studies show that the average amount of Electronically Stored Information (ESI) within a company is constantly increasing. McKinsey found that 15 out of 17 industry sectors in the United States have more data stored per company than the U.S. Library of Congress (which had 235 terabytes of information at the time of McKinsey’s study) and that companies in all sectors have at least 100 terabytes stored. Research institute IDC projects that the digital universe will reach 40ZB (40 billion TB) by 2020, exceeding previous forecasts by 5ZB. IDC noted a 20% increase in organizations that reported average ESI collection volumes of 2TB or more per matter in 2012 compared to 2010.

---

1 McKinsey, “Big data: The next frontier for innovation, competition, and productivity” (May 2011, J. Manyika, M. Chui, B. Brown et al)
3 ZB stands for Zettabyte, 1ZB = 1 billion TB (terabyte)
Regardless of the stage in the eDiscovery process, reduction of data volume has great impact and if the information is handled properly, will lead to lower costs in the next stage of the eDiscovery process. Organizations that have organized their data in a defensible way, will be able to manage their eDiscovery process in a much more effective and efficient way.

Organizations that have no or only partly control over their information, are confronted with higher eDiscovery costs then needed.
As over 90% of all cases settle prior to trial, many organizations are spending significant efforts on a legal case to estimate the risks (cost in time and money) prior to going to court. This process is called Early Case Assessment (ECA) and drastically reduces costs and expenses, provides strategic planning, optimize resources, and helps to manage budgets.

ECA helps organizations strategize against the opposing party that more often than not (especially in the United States) will try to make it as difficult as possible to comply with the discovery process.

ECA solutions: functionalities and benefits

The Early Case Assessment lifecycle will typically include all of the following:

- Perform a risk-benefit analysis.
- Place and manage a legal hold on potentially responsive documents (paper and ESI) in appropriate countries.
- Preserve information abroad.
- Gather relevant information for attorney and expert document review.
- Process potentially relevant information for purposes of filtering, search term, or data analytics.
- Information hosting for attorney and expert document review, commenting, redaction.
- Produce documents to parties in the case.
- Reuse information in future cases.

When considering solutions for ECA it is important to select software that enables you to conduct thorough ECA on an unfiltered data sample or in-place data sources as soon as you learn of a possible litigation. Functionalities should include:

- Domain analysis and pre-collection audits of data.
- Inventories of potential data matches, file count, volume, custodians, and in-network paths.
- Integration of ECA and eDiscovery platform with records management and compliance systems.
- Machine translation to address content in foreign languages.
- Crawl Search to index a selected source on-the-fly.
- Auto pre-screening “in the wild” for key issues appearing in network files and metadata.
- Metadata search to pinpoint date ranges, file types, authors.
- Concept and pattern extraction.
- First pass review.
- Graphical visualization to reveal connections among data points and custodians.

Having these functionalities will provide the following benefits:

- Quickly assess the prevalence of keywords, issues and potential liabilities within data.
- Estimate the potential scope of an impending legal hold and collection.
- Acquire metrics for a full cost-benefit analysis.
- Identify themes and patterns that can shape the strategy.
- Gain better insight to steer settlement conferences.
- Pinpoint weaknesses in internal systems that warrant future attention.

Preparing for ECA: managing e-mail

Most organizations realize that they need to take a pro-active approach to potential litigation. That also includes preparing for efficient ECA. As e-mail is considered one of the most valuable archives in litigation for discovery, (cost) savings can be realized immediately by making sure that your company’s e-mail is actively managed and archived so that information is available for ECA. An effective e-mail archiving solution helps you to make your information accessible and offers the following benefits:

- Reduced usage of high availability storage for production e-mail.
- Reduced unbudgeted mandates for staff usage during emergency collections.
- Automatic disposition according to retention rules (including legal hold).
- Early scheduling conferences compel counsel to understand the impact of search terms and other limiting criteria, and to offer up in phased discovery, the appropriate custodians.
- Increasingly, counsel also need to be aware of the value of the suit, and the budget to pursue for proportionality purposes.
- Reduced bill backs for staff time to collect or for storage to preserve. Lower risk that the archives, which are often the preservation mechanism, become obsolete, or require costly migration, in case of vendor End of Life.
- Reduced amount of data to be sent for processing by the GB.
- Reduced amount for legal review.
- Reuse of work product across cases. This allows at minimum coding of junk, Personally Identifiable Information (PII) and privilege data to be done only once.
- Control of eDiscovery data to allow data disposition in accordance with retention policies to reduce the data set for subsequent litigations.

---

5 ZyLAB, “E-mail: The Most Litigation Risky Archive”
6 One company in the Rand Study on eDiscovery costs reported that one-third of its IT department’s e-mail resources were now dedicated to preserved information. (http://www.rand.org/content/dam/rand/rand/pubs/monographs/2012/RAND_MG1208.pdf)
7 Rand study: .17 of each eDiscovery dollar was spent on processing costs.
8 Rand study: .73 of every dollar spent on electronic production in our set of 57 cases was spent on review.
Gathering and collecting information

Early in the eDiscovery process, all relevant information needs to be gathered and collected. This task often involves IT, whose expertise is required to get data out of certain systems and ready for processing for investigation and review.

In many situations, IT is tasked with converting mailboxes on e-mail servers into PST files, getting data out of document and content management systems and finding relevant information on the network. As an example; converting 20 mailboxes of each 10 GB on your Exchange Server into PST is very time consuming and after that the data still has to be processed into searchable information.

Solutions that offer various collectors to directly ingest relevant information into a review platform, are huge time savers as the system only requires the names from the custodians and it will automatically extract all e-mail messages from their mailboxes.

Handling e-mail

E-mail is the most requested format of information in eDiscovery. But e-mail is a complex format that possesses relations, duplicates and various types of attachment. E-mail can reside in different storage systems or applications. E-mail can be found on mail servers such as Exchange Server or Lotus Notes Domino, it can be found in the cloud on Office 365 or Google Business Apps, and it can be stored in personal data files, such as Outlook PST, Lotus Notes NSF, or in web-based e-mail systems such as Yahoo, Hotmail or Gmail. All these different systems require their own way of accessing and. An automated system that can comply with these different requests is a huge time saver because it allows you to:

- Collect e-mail from various sources within one application, no additional applications or manual collections are needed.
- Collect information directly from the source to keep a full trial of evidence and maximize the speed of data collections without additional (costly) steps in between.
- Perform targeted collections based on content or time to limit the total volume of data that needs to be processed and reviewed. Saving data upfront will save you lots of money in the end.
- Tap in directly to your company’s information and content repositories to collect faster and with less effort.

For organizations that are considering long term archiving of their e-mail should not only consider storage costs but also evaluate the future migration costs of their e-mail archives. Migrations are very costly, and typically cost as much as 1 year of operation of the storage system9. In ZyLAB’s white paper, “E-mail Archiving – Your First Step Towards Enterprise Information Archiving”, ZyLAB’s Enterprise Counsel, Mary Mack, reviews current eDiscovery expectations, practices and recommendations for e-mail archiving as the first step in Enterprise Information Archiving.

---

9 George Crump, the TCO of Storage, Storage Switzerland: http://www.storage-switzerland.com/Articles/Entries/2012/1/9_The_TCO_Problem_of_Storage.html
Processing

Most eDiscovery systems have search functionalities to assist reviewers and investigators. Before you can use this functionality, you need to make sure that all available information is searchable. Nothing more frustrating than knowing there is information somewhere, it is not to be found.

Electronic Files

Imagine you have a scanned contract in PDF format that ends up on the file server or on your hard disk. Using standard search tools provided by the operating system will not allow you to search the contents of the document. You may find the PDF by searching on the file name or date, but finding the contract by a contract party name will not work. And there are more formats that the standard operating system cannot search straight away; contents of compressed files such as .zip or .rar, bitmaps that contain textual information or even compound formats such as e-mail that are basically a combined set of documents, such as an e-mail file that includes various attachments. It is crucial that all information is made searchable to find all the information.

![Automated Reduction of Data Set](image)

*Figure 3 - Sample of an automated reduction of a data set during the processing stage*
Paper

Although the focus for eDiscovery lies on e-mail and electronic file formats, it is surprising to see that paper still plays a large role. After years of decrease of paper records, the number of paper records is increasing in 42% of organizations, and decreasing in 34% which results in a negative gap of 8% in 2013. When considering solutions for processing it is important to select a solution that is also able to effectively handle paper based documents by converting these into searchable information. Paper records may just have that information that is of paramount importance.

Handling PDF

PDF Normal, also known as True PDF and Real PDF, represent the ideal PDF files for most applications. These documents have been created and published using PDF software. The content includes the original formatted text of the document. Tables in the document are also usually published as formatted text. Graphics or pictures will usually appear as cut images inserted into the text.

The PDF Image is also called the Wrapped PDF or the PDF Wrapped TIFF. In these files, the content is simply an image file. The image file could be in many formats (GIF, TIFF, JPG, etc.), and of many subjects (scanned page, picture, graphic design). The most common use is a scanned page in TIFF format.

PDF Image + Text is a file type that represents a compromise between PDF Normal and PDF Image files. To make these files, the author begins with a hardcopy document. The document is scanned to get a TIFF image making it similar to the PDF Image document described above. The scan is then run through Optical Character Recognition (OCR) software to capture the text of document and the position of the text on the page. The text information is then added to the content part of the file.

Regardless of the type of electronic files or image you need to search, when selecting a solution, functionalities should include:

- Advanced OCR technology to make information searchable even when information is scanned as image or Image PDF.
- Multiple direction OCR to be able to search text in all directions.
- Language support for international languages to make sure that cross border litigation is supported.
- Proper PDF processing so that all information in PDF files is searchable, even if the information is stored in images that were part of a Word document that was published as a PDF format.
- Support for compressed file formats such as .zip, .rar or .arj.
- Identification and processing of bitmaps files such .jpg, .tif or .png and scans, pictures or photos using OCR to find relevant information later on.

10 AIIM, “Information Governance – records, risks, and retention in litigation age”, page 5
Handling of embedded objects which are file formats containing other files, for example a PDF in a Word document. Embedded objects can be important documents and by automating the processing of these embedded objects, you may find results faster.

Direct support of forensic images that requires less handling and processing activities. Files from Forensic Images don’t have to be selected manually, but these can be processed right away.

Support for different e-mail formats so that no pre-conversion has to be done. E-mails should be processed including all the attachments to an unlimited depth to ensure the highest possible retrieval capabilities.

Your system should be able to process the required volume. When needed, additional processing power should be added to ensure that data processing is done within the required time-frame. If the data volume is higher than anticipated, additional processing power can be added on the fly without reconfiguring your system set up.

Using deduplication saves on additional costs in storage and review.
Content Analysis

Content Analysis is used to understand the circumstances, facts and potential evidence in a litigation or investigation and is used to find specific information, but also to dismiss irrelevant information. The latest can be a huge time saver.

With the EDRM Doc Review Calculator\(^\text{11}\) you can calculate your savings on linear review costs. By making your data set more relevant by dismissing information and focusing on creating a relevant document set, you improve the quality of your initial data set. Lowering the number of documents to review pays off immediately; using the standard metrics in the EDRM Doc Review Calculator, reviewing 5,000,000 pages instead of 10,000,000 will generate a saving of USD 290,000.

Full-text search

Not all search is equal. Search the internet using Google, and a simple search query will instantly provide an overview of most relevant hits. In an eDiscovery however, you want to find all relevant data. Manual search across file shares, e-mail and physical records (as 40-60% of today’s companies according to research done by AIIM still use as an eDiscovery mechanism\(^\text{12}\)) is not necessary. Full-text search capabilities ease the access to information, and an eDiscovery solution like ZyLAB’s offers users instant access to the contents of documents and allows users to quickly find relevant documents. ZyLAB supports over 700 file formats, which will lower the amount of files that have to be screened manually. Advanced search capabilities are available to increase the recall and precision of your queries saving time during review.

Automated Classification

Text Mining is important to find specific information that is important for analysis or review. Various levels of text mining and classification are supported by ZyLAB’s eDiscovery solution. Text mining and classification allow the system to create sets of documents about a particular topic that need to be reviewed by a domain expert, require priority in the review process, contain privileged information or can be excluded from the review set because of obvious non-responsive information.

Search Based Classification

By using advanced keyword searches documents can be labeled or tagged. It can be used to determine documents that match a set of keywords that need to be reviewed. Keyword searches can be defined to find privileged information or

---


\(^{12}\) AIIM, “Information Governance – records, risks, and retention in litigation age”, page 20
medical information. Creating keyword searches for newsletters or messages with private conversations will allow you to dismiss information from your review set.

Regular Expression

With regular expressions specific pre-defined expressions can be defined to identify social security numbers, credit card numbers, and other expressions that need to be detected.

Entity Extraction

Entity extraction will automatically identify certain types of entities such as persons, organizations, locations, companies and many more. Using the extracted entities, documents can be grouped making the review process more efficient.

Figure 4 - Samples of searches to perform document tagging based on advanced searches

Figure 5 - Extracted Organizations from the documents can be used for further analysis
**Pattern Detection**

Using language technology specific patterns can be used to identify potential responsive information. Depending on the matter or investigation specific rules can be defined to detect patterns in the document text, indicating money transfers, travel activities, problems about products, requests for information, etc. This will support the investigative process, as it helps reviewers and investigators to quickly find relevant documents that need further investigation.

*Table 1 - Samples of text patterns that can be found in documents*

<table>
<thead>
<tr>
<th>Patterns to Detect</th>
<th>Matching Patterns Found in Data</th>
</tr>
</thead>
</table>
| **Person offer|receives trip|meeting Country** | “We are pleased to offer you a complementary trip to Thailand, where we will have more than enough time for leisure activities such as golf.”  
“Will you join us for a sailing trip to Spain?” |
| **(Organization or Person or Company) offer buys|purchase|receives from (Organization or Person or Company)** | “How much coffee, did you offered them.”  
“Can we not offer them 15 coffee?”  
“Did they accept the offer for the special compensation we discussed on the phone yesterday?” |
| **Replaced of old contracts by new contracts with different terms** | *(Replace* or supersedes) w/5 contract |

**Classification of photos and images**

A lot of information in eDiscovery and investigations is visual. Image and photo based files can be a large part of your data set. Especially e-mails contain a lot pictures. ZyLAB’s Visual Classification automatically recognizes the content of pictures and videos and identifies concepts like people, babies, elderly people, flowers, cars, planes, indoor and outdoor scenes, and many others.
The new functionality can be used for the identification of images that contain personal identifiable information (PII), potential intellectual property, handwritten notes, checks, ID’s, and other information that otherwise cannot be recognized automatically and would require full manual review. It can also be used to dismiss all holiday and party pictures from your review set.

If you want 100% recall in your system you might want to consider to OCR all bitmap based (.jpg, .png, etc.) information. This can however be a time consuming task. Therefore by classifying your bitmap based information you can select later which bitmaps need to be processed by OCR by selecting documents that are classified as Letters, spreadsheet tables, contain typed text. This will dramatically save you on processing time and will start your review process faster.

“ZyLAB has located an additional set of more than 60 previously undetected items in the EDRM Enron PST Data Set that contain explicit content, privacy, health and financial information. Responding to the invitation of EDRM, the leading standards organization for the eDiscovery and information governance market, to assist with an ongoing effort to cleanse the Enron data set, ZyLAB has shared this information with EDRM.”
Benefits and savings of classification of photos and images:

- Dramatically reduced review time.
- Almost instant accessibility when searching for PII, including ID cards, passports and driving licenses.
- Find and tag confidential information such as (photo) copies of credit cards in seconds.
- Exclude irrelevant images from the review extremely quickly. All images that contain animals, babies or cars may not be of interest for your case.
- Instant identification of all images that contain handwritten notes for further investigation.
- Reduced processing times (Performing OCR on images can be a time-consuming task, especially when you perform a four-way OCR).

Reviewing foreign languages and machine translation

Information in other languages than English can rapidly increase the cost of your eDiscovery process. Finding and hiring native speakers with the right qualifications can be difficult and costly. Besides that quality control of the review process is more difficult as this will also require additional native speakers. A cost effective alternative is using an automated machine translation solution to translate all foreign language documents to English or translate certain documents on demand during the review process.

ZyLAB’s systems leverage translation software that is based on statistics instead of complex linguistic rules. This software learns to accurately translate new information by examining previous human translations. While the machine translations may not be legally submissible in court, they do provide great insights in the content of large document and e-mail collections.
In terms of cost, Machine Translation has a significant cost advantage especially when the volumes of information that require translation are high. Machine Translation can be around 5% of what you normally pay for a professional translation\(^3\).

Besides the immediate cost benefit, the turnaround time for translations is almost instant and allows non-native reviewers to search and review the translations.

![Cost comparison between manual translation and automated machine translation](image)

**Figure 8 - Cost comparison between manual translation and automated machine translation**

**Benefits and savings of Automated Machine Translation**

- Immediate cost advantage over manual translations and fast turnaround time of translations.
- Ability to translate all documents during the processing stage for maximizing recall and access all documents in English.
- Ability to batch process documents that require translations based on native language queries which will lower the total volume of data needed for translation.
- Ad-hoc translations can be performed, when reviewers view a document that requires translation, they can submit the document for translation and review the translated document.
- Translations are linked to their original documents in the system which allows for transparent review of both original and translated version of the document.

**Searching and reviewing Audio files**

Consider the infamous audio recordings of Enron investment advisors and how clearly they demonstrated intent. This type of influential audio evidence may abound within enterprises, particularly as more data is shared (and recorded) from VOIP, Skype, MSN Live, video conferencing, unified messaging, and more. FRCP 34(a) allows sound recordings to be discoverable.

\(^3\) Translation cost comparison is based on information from Translated.net and the Machine Translation service in ZyLAB’s eDiscovery on-Demand Saas solution.
during litigation, and regulations, such as those governing investment transactions, mandate that some calls are recorded and kept readily accessible. But how can reviewers reasonably access and sift through all of this audio data?

ZyLAB eDiscovery solution offers phonetic search technology to enable users to inventory all audio files and search, review and analyze their content and metadata. Levering our phonetic search technology, our clients apply iterative search, file based tagging and dataset filtering to get maximum value from large volumes of audio data and then export it for use in downstream tools.

Human listening and speech-to-text can become cost prohibitive due to the expense of human reviewers and heavy processing drag, respectively. In fact, industry research estimates that the cost to process legacy audio data may be 80% higher than the cost to process e-mail evidence. With ZyLAB, 8 hours of audio data, such as voice-mails, can be phonetically searched in less than 1 second. The powerful phonetic search will reduce countless hours of audio data to just the potentially relevant segments that are worthy of human review.

Benefits and savings of Audio Search

- Search audio up to 350,000 times faster
- Save on your transcription costs and manual review costs. Transcriptions can cost between USD 1.00 USD and USD 2.50\(^{14}\) per minute depending on the turnaround time you require. Imagine having to transcribe 500 hours of audio, the cost of that alone are at least USD 30,000.
- Index 200 hours of recorded audio in an hour which leads to having quickly search access to your data.

![Cost of Reviewing Audio](image)

*Figure 9 - Cost comparison between reviewing audio manually and using search techniques to find relevant audio fragments*

\(^{14}\) Transcription pricing is taken from castingwords.com.
Review

Document review is a critical component of litigation and is used to identify responsive documents to produce and privileged documents to withhold. During review the legal team begins to gain a greater understanding of the factual issues and based on the type of information that is found, legal strategies emerge and develop. Document review is also the most expensive part of the eDiscovery process. Experience learns that Document review will take up to 70% of the total eDiscovery costs.

Labor cost is a major cost component of the review stage. It is therefore important that the review process supports the logic and need of the user. According to the Rand study the most-expansive claims about review speed is about 100 documents per hour\(^\text{15}\). Using tools to tag complete e-mail threads, tag duplicates and document families will increase the review speed significantly. Also documents that are classified during the analysis stage with potentially privileged will be reviewed faster as the reviewer is already advised by the system on the action that needs to be taken on these documents. Based on statistics from the use of ZyLAB’s onDemand solution, documents can be reviewed with an average speed up to 800 documents per hour, using batch tagging with sampling increases this number significantly.

Figure 10 - eDiscovery Costs using Services based on experience from ZyLAB’s eDiscovery users

\(^\text{15}\) Rand, “Where the Money Goes, Understanding Litigant Expenditures for Producing Electronic Discovery”
Benefits and savings of ZyLAB’s Review Platform:

- Reviewers do not need to wait for the complete processing of data to finish, but can start directly when the processing of data begins.
- Easy deployment of the review application to inside and outside council.
- Intuitive interface for review so that reviewers can start quickly without extensive training.
- Fast viewing of all documents without delay.
- Ability to modify the review interface to the needs of the individual reviewer to maximize performance.
- Option to tag in one go duplicates, document families, individual conversation threads, complete conversation threads.
- Ability to highlight important keywords to quickly identify important parts of documents and alert reviewers.
- Ability to create ad-hoc redactions and bulk redactions based on queries.
- The Bulk Tagging option including a sampling process will speed up the review of similar sets of documents.
- Using the Batch Assignment tool, the review load can be divided over the available reviewers more efficient.

Figure 11 - Cost comparison between a manual review and a review using ZyLAB review & analysis
The massive expansion of regulations by governments and industry groups creates an environment of increased financial penalties for non-compliance, enhanced surveillance, and damage to the reputation of non-compliant enterprises. These developments require enterprises to build highly effective programs for assuring regulatory and standard compliance.

Regulations, Compliance, Policies and Retention

Multinational organizations have to comply with different regional privacy and data protection laws, regulations and policies designed to protect individuals’ sensitive and confidential information. Compliance requires organizations to adopt and implement a variety of costly activities related to process, people and technologies.

In a benchmark study\(^{16}\) of multinational organizations the average cost of compliance for organizations is about $3.5 million. However the costs of non-compliance (fail to comply with rules, regulations, policies, contacts and other legal obligations) is much greater and is nearly $9.4 million.

In the United States, the Federal Rules of Civil Procedure (FRCP) state how litigation is carried out in U.S. federal courts. The FRCP lays out a set of rules about how electronically stored information (ESI), including e-mail needs to be handled for litigation purposes. Currently, these rules state that when a company fails to produce responsive ESI or has failed to preserve or archive relevant data, the company can be submitted to severe sanctions.

One important US regulator is the Financial Industry Regulatory Authority (FINRA), a private corporation that acts as a self-regulatory organization. It has a sharp focus on financial organizations failing to properly manage their e-mail archives and issues severe fines to firms for not properly monitoring and capturing their e-mail traffic.

- In May 2010, “FINRA Fines Piper Jaffray $700,000 for E-mail Retention Violations. The firm failed to disclose that E-mail Retention Deficiencies impacted its Ability to Respond Fully in FINRA Investigations.”\(^{17}\).
- In February 2013, fines for breaches under Health Insurance Portability and Accountability Act (HIPAA) were amended into increasing gradient from “Did not know” ($100 to $50,000 per incident) to “Willful neglect – Not corrected” ($50,000 to $1,500,000 per incident).
- In February 2013, “FINRA Fines Five ING Firms $1.2 Million for E-mail Retention and Review Violations.”\(^{18}\).
- In May 2013, FINRA accused LPL of failing to update its unwieldy and increasingly complex e-mail systems despite being “well aware” of their inability to keep up with its rapid growth.\(^{19}\).
- In June 2013, the Dutch Government proposed an amendment for the Dutch Data Protection Act that would require public Breach notification and a fine for €450,000 for a failure to notify.

---

16 Ponemon Institute, The true cost of compliance
17 http://www.finra.org/newsroom/newsreleases/2010/p121506
18 http://www.finra.org/newsroom/newsreleases/2013/p207604
19 http://www.finra.org/newsroom/newsreleases/2013/p264524
File Analysis

In order to reduce risks, analyzing what type of unstructured data that is available in your organization is essential. According to Gartner\(^{20}\) unstructured data growth is rapidly outpacing structured data and is poorly controlled and managed on file share, on personal devices and in the cloud. Organizations are poorly aware of the volume, composition, risk and business value of their unstructured data. Instead of addressing the holistic picture of unstructured data, including content, data access, and data storage, IT leaders tend to view unstructured data only from the perspective of age and do little if anything to support information governance.

"A manufacturing company had the objective to cleanse a file share environment that contained 30TB of file system data. Initially delayed because of the newness of the FA approach internally. Once permissions were received, the file discovery and analysis project took less than three months to complete. A total of 50% more content was identified beyond the original 30TB. After analysis, almost 60% of the data was identified for removal. As a result, the CIO authorized policies for the deletion of the data (currently being implemented). The ROI (payback) is two years, not including the resultant cost avoidance deferral of the storage hardware purchase."

Gartner Innovation Insight, “File Analysis Innovation Delivers an Understanding of Unstructured Dark Data”, page 4

Benefits and savings of File Analysis:

- Reduce risks by identifying which files reside where, allowing remediation on areas such as eliminating personally identifiable information, corralling and controlling intellectual property, and finding and eliminating redundant and outdated data that may lead to business difficulties such as multiple copies of a contract.
- Reducing cost by reducing the amount of data stored.
- Classify valuable business data so that it can be more easily found and leveraged.
- Supporting eDiscovery efforts for legal and regulatory investigations.

\(^{20}\) Gartner Innovation Insight, “File Analysis Innovation Delivers an Understanding of Unstructured Dark Data”
Summary

Although few companies keep track of the total costs involved in eDiscovery projects, it is safe to say that eDiscovery projects are expensive and the use of the right technology and proper information management policies can bring down these costs significantly. For each stage of the EDRM, ZyLAB can help organizations to lower their eDiscovery costs by selecting the proper technology that in combination with the right business process makes a cost effective solution. In Table 1 below, an overview is given of the savings that can be achieved by bringing an eDiscovery solution in-house and prepare your litigation risky information for eDiscovery requests.

<table>
<thead>
<tr>
<th>EDRM Stages</th>
<th>Case Size 300GB</th>
<th>Case Size 1TB</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification &amp; Preservation</td>
<td>50-70%</td>
<td>50-80%</td>
<td>-</td>
</tr>
<tr>
<td>Collection</td>
<td>30-40%</td>
<td>40-60%</td>
<td>-</td>
</tr>
<tr>
<td>Processing</td>
<td>10-20%</td>
<td>20-40%</td>
<td>-</td>
</tr>
<tr>
<td>Review &amp; Analysis</td>
<td>60-70%</td>
<td>60-85%</td>
<td>-</td>
</tr>
<tr>
<td>Production</td>
<td>10-15%</td>
<td>10-20%</td>
<td>-</td>
</tr>
</tbody>
</table>

- Using ZyLAB’s Information Risk Management solutions, organizations are eDiscovery Ready. In combination with the right information management policies and the ZyLAB software, litigation sensitive information such as e-mail is readily available for early case assessment and immediate use in the eDiscovery project.
- Collecting your information from various sources will take time. Costly migration to get information out of your systems can be prevented when your onPremise eDiscovery system can read your information management systems directly.
- It is vital to find all information that you need, not finding the important information is not an option. The integrated processing including recursive extraction of documents, deduplication, OCR, indexing over 700 file formats and much more allows a thorough analysis and review.
- Navigation tools such as e-mail threads, near duplicates, and document families increase the review speed drastically.
- Several tools support technology assisted review to classify documents, images and photos to decrease the number of documents that need to reviewed manually.
- Save on translation costs in multilingual cases.
- Use advanced search techniques to find relevant information fast.
- The integrated production system allows users to quickly select and produce documents in any format including redactions.
Important to understand is the fact that the outcome of each stage in the EDRM has a high impact on the costs of the next stage. Less is more, less data output in each stage result in more cost savings in the rest of the eDiscovery process. So if you prepare well for (potential) eDiscovery requests, then the return on investment for your onPremise eDiscovery solution will be achieved sooner than you think. Besides the economic advantages of being prepared and have the ability to run your eDiscovery project efficient, other factors such as time savings and completeness of your eDiscovery investigation must be considered as well.
## Table 1 - Average ZyLAB cost savings per eDiscovery project compared to average market pricing for stages of the EDRM

<table>
<thead>
<tr>
<th>EDRM Stages</th>
<th>Case Size</th>
<th>Case Size</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300GB</td>
<td>1TB</td>
<td></td>
</tr>
<tr>
<td>Identification &amp;</td>
<td>50-70%</td>
<td>50-80%</td>
<td>• Using ZyLAB’s Information Risk Management solutions, organizations are eDiscovery Ready. In combination with the right information management policies and the ZyLAB software, litigation sensitive information such as email is readily available for early case assessment and immediate use in the eDiscovery project.</td>
</tr>
<tr>
<td>Preservation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collection</td>
<td>30-40%</td>
<td>40-60%</td>
<td>• Collecting your information from various sources will take time. Costly migration to get information out of your systems can be prevented when your onPremise eDiscovery system can read your information management systems directly</td>
</tr>
<tr>
<td>Processing</td>
<td>10-20%</td>
<td>20-40%</td>
<td>• It is vital to find all information that you need, not finding the important information is not an option. The integrated processing including recursive extraction of documents, deduplication, OCR, indexing over 700 file formats and much more allows a thorough analysis and review.</td>
</tr>
<tr>
<td>Review &amp; Analysis</td>
<td>60-70%</td>
<td>60-85%</td>
<td>• Navigation tools such as e-mail threads, near duplicates, and document families increase the review speed drastically</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Several tools support technology assisted review to classify documents, images and photos to decrease the number of documents that need to be reviewed manually</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Save on translation costs in multilingual cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Use advanced search techniques to find relevant information fast</td>
</tr>
<tr>
<td>Production</td>
<td>10-15%</td>
<td>10-20%</td>
<td>• The integrated production system allows users to quickly select and produce documents in any format including redactions</td>
</tr>
</tbody>
</table>
## Table 2- Overview of eDiscovery costs based on research and publicly available information

<table>
<thead>
<tr>
<th>EDRM stages</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Law Technology News&lt;sup&gt;21&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>AIIIM&lt;sup&gt;22&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Minnesota Journal of Law, Science and Technology&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Rand&lt;sup&gt;24&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>GABRIEL TECHNOLOGIES CORPORATION and Trace Technologies, LLC, Plaintiff, v. QUALCOMM INCORPORATED, Snaptrack, INC. and Norman Krasner, Defendants.&lt;sup&gt;25&lt;/sup&gt;</td>
</tr>
<tr>
<td>Identification &amp; Preservation</td>
<td>$400/hour for forensic collection</td>
</tr>
<tr>
<td></td>
<td>$5,000 to $30,000/GB</td>
</tr>
<tr>
<td></td>
<td>$18,000/GB</td>
</tr>
<tr>
<td>Collection</td>
<td>$50/GB for data extraction</td>
</tr>
<tr>
<td>Processing</td>
<td>$350/GB for migration to the review platform</td>
</tr>
<tr>
<td>Review &amp; Analysis</td>
<td>$0.06/doc for technology assisted review</td>
</tr>
<tr>
<td></td>
<td>$75/hour for manual review</td>
</tr>
<tr>
<td></td>
<td>$3500/GB</td>
</tr>
<tr>
<td>Production</td>
<td>$.03/page for .tiff conversion</td>
</tr>
</tbody>
</table>

22 EDiscovery costs: Pay now or pay later, http://www.insidecounsel.com/2012/05/23/eDiscovery-costs-pay-now-or-pay-later
About ZyLAB

ZyLAB’s industry-leading, modular eDiscovery and enterprise Information Risk Management solutions enable organizations to manage boundless amounts of enterprise data in any format and language, to mitigate risk, reduce costs, investigate matters and elicit business productivity and intelligence.

For almost 30 years ZyLAB has been a dominant player in compliance and eDiscovery-related solutions, due in part to its’ advanced capabilities for multi-language support, searching, content analytics, document reviewing, and e-mail and records management (for both scanned and electronic documents).

While the ZyLAB eDiscovery & Production system is generally implemented to investigate a specific legal matter, it is a solid and robust foundation from which to pursue proactive, enterprise-wide objectives for information management. Those broader goals are achieved through the use of the ZyLAB Compliance & Litigation Readiness system.

The ZyLAB eDiscovery system is directly aligned with the Electronic Discovery Reference Model (EDRM) and features modules for forensic sound collection, culling, advanced e-mail conversion (Exchange and Lotus Notes) and legal review.

The company’s products and services are used on an enterprise level by corporations, government agencies, courts, and law firms, as well as on specific projects for legal services, auditing, and accounting providers. ZyLAB systems are also available in a Software-as-a-Services (SaaS) model.

ZyLAB’s products are extremely open and scalable, with installations managing some of the largest collections of mission-critical data in the world. The award-winning ZyLAB Information Management Platform brings our core capabilities into a single solution that provides an optimal framework for six, specialized, all-in-one system deployments.

Currently the company has sold 1.7 million user licenses through more than 9,000 installations. All of our solutions include full installation, project management and integration services. Current customers include The White House, Amtrak and US Army OIGs, US Department of Treasury, The EPA, National Agriculture Library, and Royal Library of the Netherlands, FBI, Arkansas and Ohio state police forces, German customs police, and Danish national police, War Crimes Tribunals for Rwanda, Cambodia, and the former Yugoslavia, KPMG, PricewaterhouseCoopers, and Deloitte, Akzo Nobel, Sara Lee, Pacific Life, Siemens, Dow Automotive and Lloyds of London.

ZyLAB is positioned by Gartner, Inc. as one of the strongest “Visionaries” in the 2013 Magic Quadrant for eDiscovery Software and has received numerous other industry accolades over the last 3 decennia.

ZyLAB is certified and registered as compliant with the International Standards Organization (ISO) 9001:2000. ZyLAB also lets Microsoft, Oracle and other infrastructure providers regularly certify critical components that work closely with their infrastructure. ZyLAB was certified under the US-DoD 5015.2 records management standard and ZyLAB is compliant with the European MoReq2 standard and various other regulations.

Headquartered in Amsterdam, the Netherlands and McLean, Virginia, ZyLAB also serves local markets from regional offices in New York, Barcelona, Frankfurt, London, Paris, and Singapore. To learn more about ZyLAB visit www.zylab.com.
COPYRIGHT
No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of ZyLAB Technologies B.V. The information contained in this document is subject to change without notice. ZyLAB Technologies B.V. assumes no responsibility for any errors that may appear. All computer software programs, including but not limited to microcode, described in this document are furnished under a license, and may be used or copied only in accordance with the terms of such license. ZyLAB Technologies B.V. either owns or has the right to license the computer software programs described in this document. ZyLAB Technologies B.V. retains all rights, title and interest in the computer software programs.

DISCLAIMER
This White Paper is educational in nature and is not legal advice for an organization’s particular circumstance. This White Paper is for informational purposes only. ZyLAB Technologies B.V. makes no warranties, expressed or implied, by operation of law or otherwise, relating to this document, the products or the computer software programs described herein. ZYLAB TECHNOLOGIES B.V. DISCLAIMS ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. In no event shall ZyLAB Technologies B.V. be liable for (a) incidental, indirect, special, or consequential damages or (b) any damages whatsoever resulting from the loss of use, data or profits, arising out of this document, even if advised of the possibility of such damages.

Copyright © 2009-2014 ZyLAB Technologies B.V. All rights reserved.