opentext™

Effective response to data breaches leveraging eDiscovery technology and techniques

Rapid analysis and reporting to mitigate risks

Contents

Executive summary

Managing sensitive and personal information is more complicated than ever due to rising cybersecurity threats and information volumes, a remote and hybrid workforce with "shadow IT," emerging data sources and data privacy laws.

More than 90 percent of legal operations professionals in a recent survey cited security as a major concern, up substantially from just three years ago.[1]

The likelihood of experiencing a cyber incident has increased to the point that organizations need to prepare for when—not if—an incident or data breach will occur. Yet, with newer data privacy laws and regulations, the stakes have never been higher to protect sensitive data against unauthorized access, or to notify affected parties quickly when it has been accessed.

This paper discusses legal departments' growing influence on organizational cybersecurity strategy and how eDiscovery tools and workflows can be leveraged to accelerate data breach response, analysis and reporting.

[1] Thomson Reuters, 2022 Legal Operations Index.

Legal departments' growing influence on cybersecurity strategy

Managing sensitive and personal information has always been complex. It is more complicated than ever due to the rise in cybersecurity threats, an ever-expanding pool of information, a remote/hybrid workforce with "shadow IT," new and emerging data sources and data privacy laws. In response, the legal department's influence on information security decisions has increased, and with good reason. In a recent survey, more than 90 percent of legal operations professionals cited security as a major concern, up substantially from just three years ago.[2]

The recent 2022 State of Cybersecurity Report found that 84 percent of CLOs in surveyed companies now have at least some cybersecurity-related responsibilities, with 48 percent citing responsibility for coordinating cyberlaw strategy across the entire organization.[3]

Regulatory environment

New or strengthened laws and regulations compel organizations to better document and protect the data they collect. The rules are changing rapidly and vary across regions, making it difficult for organizations to keep up. eDiscovery workflows and technology have become increasingly vital to addressing those compliance obligations.

Data breach notification is legally required (with varying conditions) within the U.S, EU and other markets. Private civil litigation is highly probable after a major data breach, as plaintiffs' counsel firms are increasingly taking advantage of these events to commence legal action. According to one report, major data breach class actions have increased by 72 percent in just two years.[4]

Complaints are typically brought within weeks after breach announcements, making proactive mitigation even more important for organizations facing potential exposure. Any data breach response program should include preparation for potential litigation.

Extending traditional eDiscovery technology and techniques to the data breach lifecycle

In the eDiscovery Today 2023 State of the Industry Report, 410 respondents identified use cases where they apply eDiscovery technology and workflows. Forty-seven percent included incident response from data breaches as a key application.[5] The key to breach response analysis and reporting is quickly and accurately searching, identifying and reporting on affected data. As a result, legal teams are turning to well-developed eDiscovery technology and workflows.

Applying eDiscovery and investigative workflows to data breach analysis and response

Following a data breach, the task of determining exactly **what** personal or confidential information is contained in the body of data taken and to **who** the information belongs can be a daunting task. eDiscovery and investigative workflows incorporate many of the same techniques and tools used in litigation to answer these questions.

[2] Thomson Reuters, 2022 Legal Operations Index.

[3] Association of Corporate Counsel (ACC) Foundation and Ernst & Young, 2022 State of Cybersecurity Report.

[4] Morrison Foerster, Privacy Litigation 2022 Year in Review: Data Breach Litigation. (2023)

[5] eDiscovery Today, 2023 State of the Industry Report. (2023)

A typical eDiscovery workflow for data breach analysis involves five phases:

1. Ingestion and processing

2.Early data analysis

3.Data mining

4.Data extraction

5.Reporting

Below are the current tools, techniques and considerations associated with each phase.

Phase 1: Ingestion and processing

Every eDiscovery workflow begins with collection of potentially responsive data, in this case, the data potentially affected by the breach. If the incident is a "smash and grab," a vast quantity of data may be exfiltrated. In those circumstances, identifying the total corpus of data compromised is only the start of the process.

Once the initial forensic team has identified the outline of the body of data that was compromised, the organization will need a deeper understanding of the content of the data that was taken, including who was identified in that data. This is key to understanding the potential risk and damage associated with the breach, whether legislative reporting obligations have been triggered, and ultimately to fulfill data reporting obligations where required. To achieve this, the compromised data needs to be processed and ingested into an eDiscovery platform.

Phase 2: Early data analysis

Meeting applicable time-bound regulatory reporting obligations to notify individuals whose personal data may have been compromised requires peak efficiency. The best way to uncover what and who is affected in a data breach incident is to leverage analytics to rapidly separate data in two broad categories—those unlikely to contain personal, sensitive or confidential data and those that require more careful scrutiny.

While the goals and workflows associated with data breach analysis and reporting are distinct from a litigation eDiscovery review for production scenario, some of the same advanced tools and analytics created for eDiscovery can be used to maximize efficiency, including rapid analytics investigative review (RAIR).[6]

In this initial phase, the RAIR approach uses analytics tools, including communication analysis tools, domain analysis and knowledge of the organization's structure and communication patterns, to isolate a much smaller data set for review by the analysis and reporting team. This enables a rapid understanding of underlying structure and patterns within the dataset to efficiently slice through large swaths of the data. RAIR can separate data that likely contains personal information from that which can be ignored, such as mundane business communications.

[6] OpenText's RAIR approach is discussed in the position paper, Maximizing document review efficiency with rapid analytic investigative review.

Further eDiscovery review techniques and technology are invaluable to reduce the volume of data needing review. These include metadata, textual and conceptual analytics. For example, sender domains may identify innocuous marketing materials and newsletters, folder or file names can identify recurrent or pro forma documents, each of which are likely to be excluded. It is important to conduct some level of sampling to confirm decision-making.

Email thread identification eliminates all but the most-inclusive email(s) for every thread. It is important to ensure text and attachment comparison, as well as exercise care in managing attachment context as part of the process. This may eliminate earlier emails in the thread with personal information, which will prevent the ability to conduct a frequency of occurrence analysis.

Phase 3: Data mining

The goal of this phase is to locate documents containing PII and PHI within the smaller culled document set. Data mining refers to discovery and extraction of patterns and knowledge from data sets of structured and unstructured data. It involves finding data subjects and elements using only automation and computer capabilities.

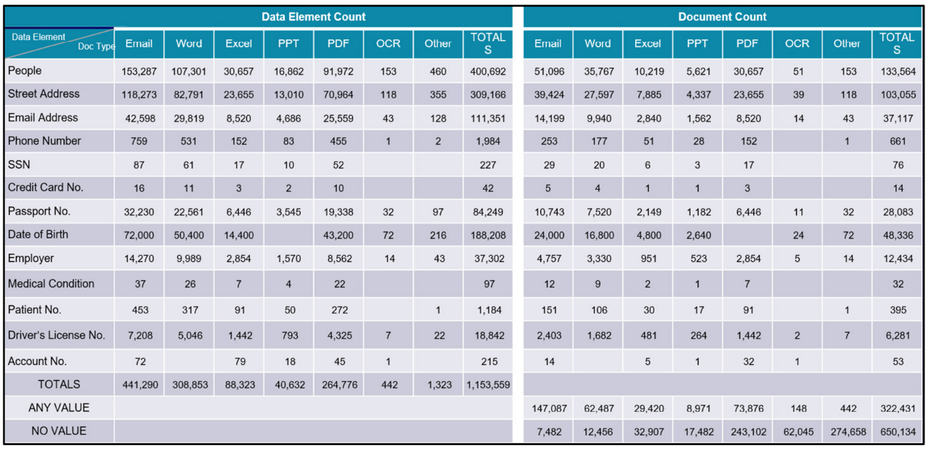
Data mining is typically performed using two approaches:

1) Pattern Matching (RegEx), which identifies strings that match specific patterns.

Pattern matching can be used for virtually any identifiable pattern, such as credit card numbers and email addresses, across either individual or thousands of documents. It can save time, labor and costs while protecting against the inadvertent exposure of confidential data. Advanced text analytics identifies occurrences of data subjects' names and can facilitate automatic redaction of that data. Automated QC processes ensure accuracy and reduce the risk of inadvertent non-compliance with data privacy laws.

2) AI and machine learning technology, which uses unsupervised learning capabilities to identify documents with potential PII or PHI.

Potential benefits of data mining include identifying documents that will likely need review through automated reporting, establishing an initial data subject list and controlling batching to send data regarding the same subject to the same reviewer (to the extent possible) to improve consistency.



	Data Element Count								Document Count							
Data Element / Doc Type	Email	Word	Excel	PPT	PDF	OCR	Other	TOTALS	Email	Word	Excel	PPT	PDF	OCR	Other	TOTALS
People	153,287	107,301	30,657	16,862	91,972	153	460	400,692	51,096	35,767	10,219	5,621	30,657	51	153	133,564
Street Address	118,273	82,791	23,655	13,010	70,964	118	355	309,166	39,424	27,597	7,885	4,337	23,655	39	118	103,055
Email Address	42,598	29,819	8,520	4,686	25,559	43	128	111,351	14,199	9,940	2,840	1,562	8,520	14	43	37,117
Phone Number	759	531	152	83	455	1	2	1,984	253	177	51	28	152		1	661
SSN	87	61	17	10	52			227	29	20	6	3	17			76
Credit Card No.	16	11	3	2	10			42	5	4	1	1	3			14
Passport No.	32,230	22,561	6,446	3,545	19,338	32	97	84,249	10,743	7,520	2,149	1,182	6,446	11	32	28,083
Date of Birth	72,000	50,400	14,400		43,200	72	216	188,208	24,000	16,800	4,800	2,640		24	72	48,336
Employer	14,270	9,989	2,854	1,570	8,562	14	43	37,302	4,757	3,330	951	523	2,854	5	14	12,434
Medical Condition	37	26	7	4	22			97	12	9	2	1	7			32
Patient No.	453	317	91	50	272		1	1,184	151	106	30	17	91		1	395
Driver's License No.	7,208	5,046	1,442	793	4,325	7	22	18,842	2,403	1,682	481	264	1,442	2	7	6,281
Account No.	72		79	18	45	1		215	14		5	1	32	1		53
TOTALS	441,290	308,853	88,323	40,632	264,776	442	1,323	1,153,559								
ANY VALUE									147,087	62,487	29,420	8,971	73,876	148	442	322,431
NO VALUE									7,482	12,456	32,907	17,482	243,102	62,045	274,658	650,134

Example of automated report created during data mining

Phase 4: Data extraction

In the extraction phase, a small data analysis and reporting team conducts a targeted review and extracts the critical details—the what and the who—for the purposes of creating the data breach report. The team again leverages textual and pattern recognition analytics to identify and extract personal information with maximum speed and accuracy.

Phase 5: Reporting

Reporting is the final step, which involves considerations for normalization, deduplication and QC. Those include:

- **Programmatic normalization**
  Ensures data value consistency (e.g., names, addresses, etc.) for deduplication and reporting.
- **Deduplication and data consolidation**
  Uses both programmatic deduplication, which relies on an exact data subject match and "unique" data elements, and limited, final pass manual deduplication.

When considering how far is "too far" when it comes to normalization and deduplication—it is often better to "over-notify" vs. "under-notify" data subjects.

Until such time as the technology and the law around data breach analysis and reporting becomes more settled, there are more questions than there are answers. A RAIR approach provides a defensible and practical way to answer what and who questions of data breach analysis for reporting.

Potential future applications of eDiscovery tools and techniques

What is on the horizon? Here are some potential future eDiscovery applications:

- Improving existing automated data extraction techniques, such as Pattern Matching (RegEx) entity extraction and address recognition.
- Developing an association capability between data subject and data elements.
- Applying continuous active learning to enhance data extraction and association and enhance continuous PII/PHI source identification.
- Developing relational database capabilities for contemporaneous data assessment.
- Incorporating AI into report QC and deduplication and consolidation.

Driving a proactive approach to cybersecurity

Lawyers can take additional leadership roles in managing cybersecurity and data risks—not just response and reporting—by driving an active process that includes:

Third-party risk management

Corporate legal can take a more active role by conducting a thorough vendor contract review to ensure that vendors and other third parties adhere to company standards regarding data security. This includes processes and procedures to protect information when transferred to third parties and requiring in contracts that third parties provide prompt notification in the event of a breach.

Taking an information assurance approach to GRC

Because of their involvement in governance, risk management and compliance (GRC), legal teams have been placed in a leadership role to ensure a program that manages and protects sensitive data.

Protecting data starts with knowing where the sensitive data is within the organization to protect. Information Assurance is the discipline of efficiently and defensibly identifying, preserving and collecting information from various organizational endpoint data sources to support key discovery business objectives. Corporate legal can and should be a leader in this process, as it drives the success of the organization's cybersecurity strategy along with other benefits, including streamlining processes such as eDiscovery.

Putting eDiscovery tools and techniques to work

Just as litigation involves an eDiscovery workflow from identification through production of potentially responsive ESI, a similar workflow can be applied to incident response. eDiscovery tools and techniques can be an extremely helpful in identifying personal data that has been potentially exposed, reviewing data and documents to identify customers to which the data exposure may apply, and notifying those customers. Corporate legal can take the lead in coordinating that workflow to rapidly support incident response for the organization.

Learn how eDiscovery workflows can be applied to address other use cases, including responding to subject rights requests, internal investigations and M&A due diligence in Powerful new applications for eDiscovery technology and techniques.

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- OpenText CEO Mark Barrenechea's blog
- Twitter | LinkedIn

opentext.com/contact