Six Important eDiscovery Trends for 2024

and How to Prepare for Them

January 2024

Contents

Introduction

In 1849, French writer Jean-Baptiste Alphonse Karr wrote the phrase "plus ça change, plus c'est la même chose." In English, that translates to: "the more things change, the more they stay the same." While we don't think that Kerr had electronic discovery in mind with that statement – after all, the emergence of electronically stored information (ESI) was still over a century away – his phrase couldn't be more perfect to describe the state of eDiscovery today. eDiscovery challenges are continually changing and evolving due to several factors, including:

- **The ESI:** In the Big Data era, data volumes in the world have skyrocketed, adding challenges in meeting discovery deadlines and budgets. Not only that, but the variety of discoverable data types has mushroomed over that time as well.
- **The Cloud:** Organizations have been migrating to cloud-based solutions for years, and that migration accelerated considerably during the pandemic, leading to a significant increase in the use of enterprise solutions in the cloud.
- **The Data Protection Challenges:** With data breaches at an all-time high, it's becoming more difficult than ever to protect personal and sensitive data. Yet, the stakes are higher than ever with ever changing data privacy laws and the fines for failing to comply with data privacy laws increasing in both frequency and amount.
- **The Technology Challenges:** The evolution of technology has moved into hyperdrive with the advent of artificial intelligence (AI) and now generative AI. While generative AI shows great promise, it's current iteration has also yielded concerns about accuracy and "hallucinations," as well as privacy and copyright protection concerns.

Challenges like these are shaping how organizations are conducting eDiscovery today and tomorrow. As Veritas, we see six important trends that will impact how eDiscovery is conducted in 2024. They are:

- **Trend 1 – Move to the Left of the EDRM:** To address data challenges and conduct eDiscovery efficiently and cost-effectively, organizations must focus more on "left-side" EDRM phases.
- **Trend 2 – Increased Variety of Data Types:** Organizations must be continually creating and refining workflows to support more data types than ever.
- **Trend 3 – Integration, Not Assimilation, with Enterprise Archival and Compliance Solutions:** Organizations must integrate with enterprise solutions to reduce manual processes to preserve and collect ESI, while taking advantage of advanced eDiscovery capabilities included in a platform designed for eDiscovery.
- **Trend 4 – Protection of Personal and Sensitive Data:** Organizations must leverage technology to identify and protect personally identifiable information (PII) before and during discovery.
- **Trend 5 – Increased Importance of Metadata in Discovery:** With a greater potential of AI-generated "deepfakes" in evidence, metadata for ESI will be more important than ever for evidence authentication.
- **Trend 6 – Leveraging AI for Proven Use Cases:** Organizations must be proactive, yet cautious in their adoption of AI technology, focusing on reliable use cases with proven benefits.

Let's explore these trends in more detail in terms of their impact on organizations, and how to prepare for them.

Trend #1: Move to the Left of the EDRM

In the era of Big Data, data volumes are soaring, and organizations are struggling to keep up. According to IDC, the global data sphere is expected to reach 175 zettabytes by 2025, with a compounded annual growth rate of 61 percent. Another projection by Statista has global data at 181 zettabytes by 2025! That's 181 trillion gigabytes! When the Electronic Discovery Reference Model (EDRM) model was created back in 2005, the global data sphere was **0.1 zettabytes**. That means global data will have grown 1,750 to 1,810 times in 20 years!

The assumed workflow to the practice of eDiscovery defined by the EDRM model – where ESI has historically moved through a progression of being identified, preserved, collected, processed, analyzed, reviewed, produced and presented – can no longer support today's data volumes. The volume of ESI and the variety of ESI sources have evolved dramatically over the years and that has forced eDiscovery to move UPSTREAM on the EDRM, with a greater focus on Early Data Assessment (EDA) at the source where the data lives, to enable more targeted collection of potentially responsive ESI. No wonder EDRM is currently assessing the EDRM model with their EDRM 2.0 Working Group and looking at upstream considerations as a part of that assessment!

The need to conduct EDA earlier than ever makes it necessary to leverage technology, including AI and machine learning technology, such as advanced filtering, sentiment analysis, and classification tags. It's important to apply those technologies to support that analysis sooner than ever, which means that information governance and eDiscovery need to be more integrated than ever. Integrating information governance, identification, preservation and collection – the "left-side" phases on the EDRM – has become vitally important to conducting eDiscovery efficiently and cost-effectively. Organizations need to conduct analysis of data — where it lives, across various sources, using the latest AI tools — to support both information governance and eDiscovery.

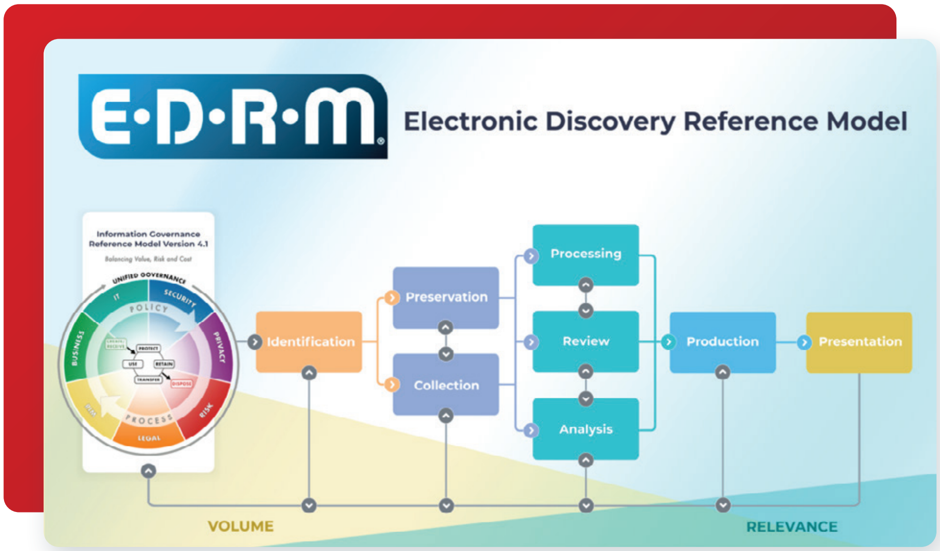


Figure 1: EDRM Model with Focus on "Left-Side" Phases (Source: EDRM)

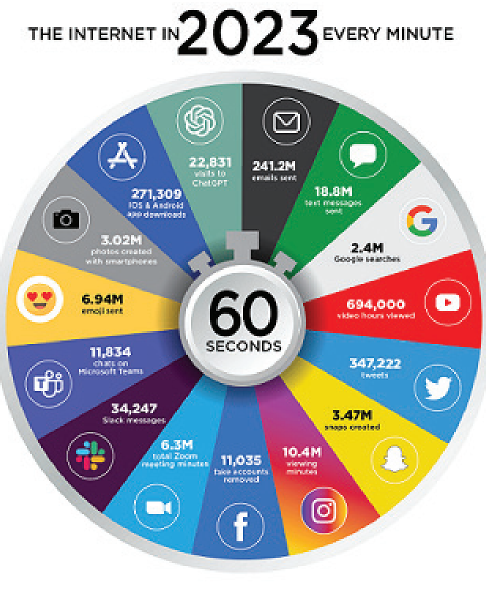
eDiscovery solutions today must include the ability to analyze, classify and control unstructured data through early data assessment where the data lives, across various data sources, using the latest proven AI techniques.

It's important to select a solution that supports your needs over the entire EDRM, beginning with Information Governance, not Identification.

Trend 2: Support an Increased Variety of Data Types

In this blog post, we identified increasing types of data as the biggest eDiscovery challenge. Since then, the Fall 2023 eDiscovery Business Confidence Survey published by ComplexDiscovery identified "Increasing Types of Data" as the biggest issue respondents felt will most impact the business of eDiscovery over the next six months (out of six options) with 37.5% of respondents, the eighth time in nine surveys it was at (or tied for) the top issue!

No wonder. The 2023 Internet Minute infographic created by eDiscovery Today and LTMG illustrates the volume and variety of data out there today in terms of what happens in an average minute on the internet. Everything from 241 million emails sent per minute to nearly 19 million text messages sent per minute to Slack messages, Teams chats, and more. Many of these are data types that organizations must address in discovery today.



Created by: eDiscovery Today & LTMG

Figure 2: 2023 Internet Minute Infographic (Source: eDiscovery Today and LTMG)

Types of data that are subject to eDiscovery and compliance today include:

- **Modern Collaboration Platforms:** Includes not only the most common collaboration platforms like Slack, Microsoft Teams and Zoom, but other collaboration platforms used by many organizations like MS Teams Channel, Salesforce Chatter, Yammer and Webex Teams.
- **Document Collaboration:** Includes collaborative, cloud-based document solutions like Google Drive, Dropbox and Microsoft OneDrive.
- **Mobile and Text:** Includes text message data from providers like AT&T and Verizon, and messaging sources like SMS, WhatsApp and WeChat, China's most popular messaging app.
- **Financial Platforms:** Includes financial platforms like Bloomberg, Symphony and Yieldbroker.
- **Other Sources:** Includes recorded / transcribed meetings, enterprise tools (e.g., IT Service Management (ITSM) tools like ServiceNow), SQL databases, JSON files and more.

If you're ignoring data sources like these, you could be violating your discovery duties, and possibly subjecting yourself to sanctions! That's why it's important to select a technology solution for compliance and eDiscovery that supports a wide range of sources to support those requirements.

Trend 3: Integration, Not Assimilation, with Enterprise Archival and Compliance Solutions

Figure 3: Veritas Alta™ Capture Supports Over 120 Data Sources (Source: Veritas)

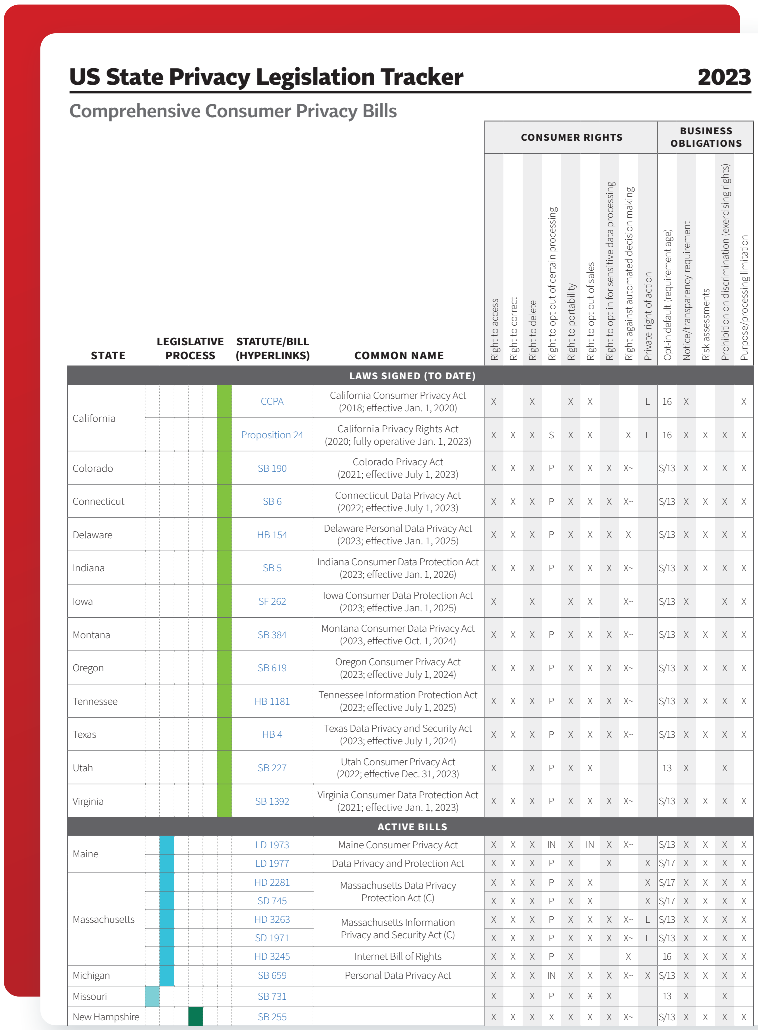
The move to the cloud means that much of the ESI discoverable today is in enterprise solutions in the cloud. One key example of the significant shift to cloud-based solutions today is Microsoft 365. Microsoft has transitioned its widely used Office applications, such as Word, Excel, PowerPoint, and Outlook, to the cloud. Additionally, M365 offers tools like SharePoint and OneDrive for document management, as well as Teams for online collaboration and meetings. And it incorporates data governance and eDiscovery features through Microsoft Purview!

Why not just use Microsoft Purview for archival, compliance and eDiscovery? Because there are several limitations you need to be aware of when it comes to using Purview that can impact your eDiscovery workflows. They include: 1) the inability to index more than 2 million characters of any document, 2) slow indexing with a maximum advanced indexing throughput is 2GB per hour, 3) slow searching with searches that take as long as 25 minutes for large organizations, 4) Purview's tendency to throttle the export of data to load balance environments, 5) limited file type support with support of only 63 file types, and 6) no in-place review and analytics with Purview – you must copy the data into a case first, which can take hours. That's why we're talking "integration", not "assimilation".

While eDiscovery starts with how your data is managed and archived for information governance purposes, it's important to have a robust solution for effectively conducting discovery on ESI from enterprise solutions.

Factors to look for include:

- **Upstream early data assessment:** The solution should be able to support conducting early data assessment at the source where the data lives.
- **Expansive data capture and collection with full conversational context:** The solution should support collection from a large range of data sources (such as the examples mentioned in Trend #2 above) and should capture communications in a manner that preserves the full context of the conversation.
- **Comprehensive indexing and advanced, iterative searching:** The solution should include advanced indexing technology and powerful search capabilities that deliver results in seconds.
- **Support collaborative workflows:** The solution should streamline matter management and support the ability for your organization's legal team and outside counsel to work together effectively.
- **Self-service capabilities:** The solution should enable designated internal and external reviewers and administrators to perform self-service online exports of search results, minimizing disruption to your organization's IT team.



US State Privacy Legislation Tracker **2023**

Comprehensive Consumer Privacy Bills

STATE	LEGISLATIVE PROCESS	STATUTE/BILL (HYPERLINKS)	COMMON NAME	CONSUMER RIGHTS									BUSINESS OBLIGATIONS				
				Right to access	Right to correct	Right to delete	Right to opt out of certain processing	Right to portability	Right to opt out of sales	Right to opt in for sensitive data processing	Right against automated decision making	Private right of action	Opt-in default (requirement age)	Notice/transparency requirement	Risk assessments	Prohibition on discrimination (exercising rights)	Purpose/processing limitation
LAWS SIGNED (TO DATE)																	
California	🟩	CCPA	California Consumer Privacy Act (2018; effective Jan. 1, 2020)	X		X		X	X			L	16	X			X
	🟩	Proposition 24	California Privacy Rights Act (2020; fully operative Jan. 1, 2023)	X	X	X	S	X	X		X	L	16	X	X	X	X
Colorado	🟩	SB 190	Colorado Privacy Act (2021; effective July 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Connecticut	🟩	SB 6	Connecticut Data Privacy Act (2022; effective July 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Delaware	🟩	HB 154	Delaware Personal Data Privacy Act (2023; effective Jan. 1, 2025)	X	X	X	P	X	X	X	X		S/13	X	X	X	X
Indiana	🟩	SB 5	Indiana Consumer Data Protection Act (2023; effective Jan. 1, 2026)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Iowa	🟩	SF 262	Iowa Consumer Data Protection Act (2023; effective Jan. 1, 2025)	X		X		X	X		X~		S/13	X		X	X
Montana	🟩	SB 384	Montana Consumer Data Privacy Act (2023, effective Oct. 1, 2024)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Oregon	🟩	SB 619	Oregon Consumer Privacy Act (2023; effective July 1, 2024)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Tennessee	🟩	HB 1181	Tennessee Information Protection Act (2023; effective July 1, 2025)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Texas	🟩	HB 4	Texas Data Privacy and Security Act (2023; effective July 1, 2024)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Utah	🟩	SB 227	Utah Consumer Privacy Act (2022; effective Dec. 31, 2023)	X		X	P	X	X				13	X		X	
Virginia	🟩	SB 1392	Virginia Consumer Data Protection Act (2021; effective Jan. 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
ACTIVE BILLS																	
Maine	🟦	LD 1973	Maine Consumer Privacy Act	X	X	X	IN	X	IN	X	X~		S/13	X	X	X	X
	🟦	LD 1977	Data Privacy and Protection Act	X	X	X	P	X		X		X	S/17	X	X	X	X
Massachusetts	🟦	HD 2281	Massachusetts Data Privacy Protection Act (C)	X	X	X	P	X	X			X	S/17	X	X	X	X
	🟦	SD 745		X	X	X	P	X	X			X	S/17	X	X	X	X
	🟦	HD 3263	Massachusetts Information Privacy and Security Act (C)	X	X	X	P	X	X	X	X~	L	S/13	X	X	X	X
	🟦	SD 1971		X	X	X	P	X	X	X	X~	L	S/13	X	X	X	X
	🟦	HD 3245	Internet Bill of Rights	X	X	X	P	X			X		16	X	X	X	X
Michigan	🟦	SB 659	Personal Data Privacy Act	X	X	X	IN	X	X	X	X~	X	S/13	X	X	X	X
Missouri	🟦	SB 731		X		X	P	X	~~X~~	X			13	X		X	
New Hampshire	🟩	SB 255		X	X	X	X	X	X	X	X~		S/13	X	X	X	X

Figure 4: US State Privacy Legislation Tracker (Source: International Association of Privacy Professionals)

Trend 4: Protection of Personal and Sensitive Data

Protecting personally identifiable information (PII) has never been more complex. Before 2023, there was only one US state that had a comprehensive data privacy law in place: California, with the California Consumer Privacy Act (CCPA) that was enacted in 2020. That law was replaced by the California Privacy Rights Act, which went into effect on January 1, 2023. Four other state data privacy laws are also going into effect in 2023. And another seven states have enacted comprehensive data privacy laws in 2023, resulting in 12 states that have enacted a law. And each law is a bit different. The US state privacy legislation tracker from IAPP shows the progression of privacy laws in the US.

Not only that, but the stakes have never been higher. The General Data Protection Regulation (GDPR) went into effect in May 2018. For the first three years, the total cumulative amount of fines only reached about a quarter of a billion euros. Over the past two years, the total cumulative amount of fines has skyrocketed, with **close to 4 billion euros in fines over that time** and there have been over 1,800 fines issued through GDPR in 5 years.

As a result, data protection is vital and industry-standard data protection mechanisms such as role-based access control (RBAC), multi-factor authentication and military-grade encryption in-transit and at rest are a "must have" to maximize protection of sensitive data and comply with data protection regulations, such as FINRA, SEC, MiFID II, GDPR, CPRA, HIPAA, PCI, ITAR, SOX, and more.

Creation of customizable auto classification rules to identify personal data, such as PII or protected health information (PHI) is key to identifying where that sensitive data is located. You can't protect what you can't find.

Figure 5: Image Generated by DALL-E 3 (Source: Microsoft)

Trend 5: Increased Importance of Metadata in Discovery

Metadata has always been important in eDiscovery – it not only enriches the information available but also enhances the efficiency and accuracy of the process. Metadata provides several key benefits, including:

- **Context:** Metadata can provide critical context for a document. For instance, metadata might show when a document was created, by whom, who else had access to it, and how often it was edited or viewed. This information can provide insights into the significance or relevance of a document.
- **Search and Organization:** Metadata allows for efficient and targeted searches, helping legal teams quickly locate relevant documents among potentially millions of files.
- **Timeline Analysis:** Metadata, such as sent and received dates, can be used to construct timelines of events, which can be vital in understanding the sequence of events in a case.
- **Deduplication:** Redundant data can result in increased costs and time spent on review. Metadata such as Hash values can help identify and remove duplicates, ensuring that legal teams are not reviewing the same document again and again.
- **Chain of Custody:** Metadata can help with chain of custody to demonstrate that electronic evidence has been handled properly and has not been altered.
- **Linking:** Some metadata, such as email thread fields, can show relationships between files or communications.
- **Assist in Privilege Review:** Metadata in emails to identify senders and recipients can help legal teams identify potentially privileged communications that might be exempt from production.

One of the downsides to the emergence of generative AI in mainstream society is that the ability to create convincing deepfakes of images, audio or video evidence has been greatly improved.

Because of this recent additional threat, the ability to authenticate your evidence has become more important than ever. Complete and accurate metadata (including full conversational context of messages) is the biggest key to ensuring proper evidence authentication today, which makes that metadata more important than ever to your discovery workflows.

Your technology solution should capture data defensively, preserving the metadata as it resides in-place at the data source. This ensures that chain-of-custody is preserved and positions your team to respond to the "deepfake defense".

Trend 6: Leveraging AI for Proven Use Cases

With the introduction of ChatGPT in late 2022, generative AI became a worldwide phenomenon in 2023. And it has shown the ability to accomplish amazing things. GPT-4, which was introduced in March, not only passed the bar exam, it scored in the 90th percentile. It also can suggest a meal based on the contents shown in a picture of the interior of a refrigerator, and even write code in popular programming languages such as JavaScript, PHP and Python! It can also generate sophisticated images in seconds based on a simple prompt, such as the one below generated by OpenAI's DALL-E 3 using the phrase "technology trends in 2024".

Generative AI has the potential to generate great benefits, but it also has the potential to generate issues. In 2023, a lawyer submitted a filing generated by ChatGPT with at least six "bogus judicial decisions with bogus quotes and bogus internal citations" that were "hallucinations" of the platform. There have also been copyright concerns and defamation claims, as well as concerns about data protection and using the technology to launch cyberattacks.

The benefits vs. the risks of generative AI technology are still unknown. But there are proven use cases for AI technology that can be applied upstream in discovery to streamline eDiscovery workflows. Two examples are:

- **Sentiment Analysis:** The use of natural language processing (NLP) to determine the attitude, sentiment, or emotion of a subject based on their content is sentiment analysis. In eDiscovery, it can be used to: 1) prioritize documents that are more likely to be relevant, 2) identify patterns in communication that may be relevant to a case, and 3) flag content with a strong negative sentiment for further review.
- **Auto Classification:** The use of advanced algorithms and machine learning models to automatically organize information into predefined classes or categories is auto classification. Example uses of auto classification include: 1) identifying PII or other sensitive data, 2) identification and remediation of Redundant, Obsolete, or Trivial (ROT) data, and 3) foreign language detection.

Until the benefits of generative AI technology have been reliably demonstrated, it's best to stick to use cases for AI that have shown to provide consistent and proven benefits in streamlining discovery workflows.

How to Prepare for These Trends

As a leader in multi-cloud management, Veritas helps ensure the protection, recoverability, and compliance of their data to over 80,000 customers. Veritas has a reputation for reliability at scale, which delivers the resilience its customers need against the disruptions threatened by cyberattacks, like ransomware. No other vendor is able to match the ability of Veritas to execute, with support for 800+ data sources, 100+ operating systems, 1,400+ storage targets, and 60+ clouds through a single, unified approach.

In order to prepare for these eDiscovery trends, we would like to introduce Veritas Alta eDiscovery!

Veritas Alta eDiscovery is a cloud-based, end-to-end eDiscovery solution that enables organizations to collect, review and produce electronically stored information for your organization's legal and investigation needs. Veritas Alta eDiscovery provides a complete discovery solution: defensible collection of your organization's relevant content sources, purpose-built review of all types of data, and efficient production of your relevant documents.

Our end-to-end eDiscovery solution supports cross-functional collaboration on matters and investigations so that your internal and external legal teams, IT, HR, and other organizational stakeholders can work more efficiently together. Our product provides these professionals with the built-in tools they need to quickly locate their intended dataset, including real-time iterative search capabilities, built-in collaborative eDiscovery workflow, and flexible SaaS export options.

Veritas Alta eDiscovery features and capabilities allow organizations to keep up with these industry trends.

Trend 1 – *Move to the Left of the EDRM: Alta eDiscovery is an end-to-end solution that helps companies manage the growth of their data and move upstream towards Advanced early data assessment on the EDRM model.*

Trend 2 – *Increased Variety of Data Types: We can capture data from over 120+ content sources allowing organizations to store multiple data sources and types.*

Trend 3 – *Integration, Not Assimilation, with Enterprise Archival and Compliance Solutions: Alta eDiscovery is an advanced SaaS solution using powerful search functionality and workflows performing automatically based on search criteria to significantly reduce the review time.*

Trend 4 – *Protection of Personal and Sensitive Data: Integrated with Alta eDiscovery, Veritas has many other products such as Classification and Surveillance which assist in confirming all latest compliance and governance regulations are being honored.*

Trend 5 – *Increased Importance of Metadata in Discovery: Veritas Alta eDiscovery stores and indexes every email and attachment, content from unified communications and instant messaging systems, and any other metadata an organization has defined important.*

Trend 6 – *Leveraging AI for Proven Use Cases: As generative AI keeps developing and growing, Veritas Alta eDiscovery expands its capabilities of collecting from any data source including ChatGPT.*

Conclusion

The society and business world we live in today is continually changing, which impacts how we conduct eDiscovery today. Organizations must be prepared to address trends in eDiscovery technology and workflows that include:

1. The growth in data volumes.
2. Increased variety of data sources.
3. The popularity of cloud-based enterprise solutions.
4. The continual change and increased enforcement of data privacy laws.
5. The emergence of deepfakes and the "deepfake defense."
6. The uncertainty of current applications of generative AI technology.

Veritas sees these six trends as important to address for organizations in 2024, and its important to select eDiscovery technology solutions which:

1. Span the entire EDRM life cycle (including Information Governance).
2. Comprehensively support a large range of file types.
3. Provide seamless integration to cloud-based enterprise solutions.
4. Automate the identification of personally identifiable information (PII) and provide industry-standard data protection mechanisms.
5. Preserve complete and accurate metadata (including full conversational context of messages).
6. Support AI use cases that have proven to provide benefits.

Change is inevitable. Organizations that embrace that change are more likely to conduct eDiscovery in a manner that is efficient and cost-effective in 2024 and beyond!

About Veritas Alta eDiscovery

Veritas Alta eDiscovery is a cloud-based, end-to-end eDiscovery solution that enables organizations to collect, review and produce electronically stored information for your organization's legal and investigation needs. Veritas Alta eDiscovery provides a complete discovery solution in one step: defensible collection of your organization's relevant content sources, purpose-built review of all types of data, and efficient production of your relevant documents.

About Veritas

Veritas Technologies is a global leader in data protection and availability. Over 80,000 customers—including 87 percent of the Fortune Global 500—rely on us to abstract IT complexity and simplify data management. The Veritas Enterprise Data Services Platform automates the protection and orchestrates the recovery of data everywhere it lives, ensures 24/7 availability of business-critical applications, and provides enterprises with the insights they need to comply with evolving data regulations. With a reputation for reliability at scale and a deployment model to fit any need, Veritas Enterprise Data Services Platform supports more than 800 different data sources, over 100 different operating systems, more than 1,400 storage targets, and more than 60 different cloud platforms. Learn more at www.veritas.com. Follow us on Twitter at @veritastechllc.

VERITAS™

2625 Augustine Drive
Santa Clara, CA 95054
+1 (866) 837 4827
veritas.com

For global contact
information visit:
veritas.com/company/contact