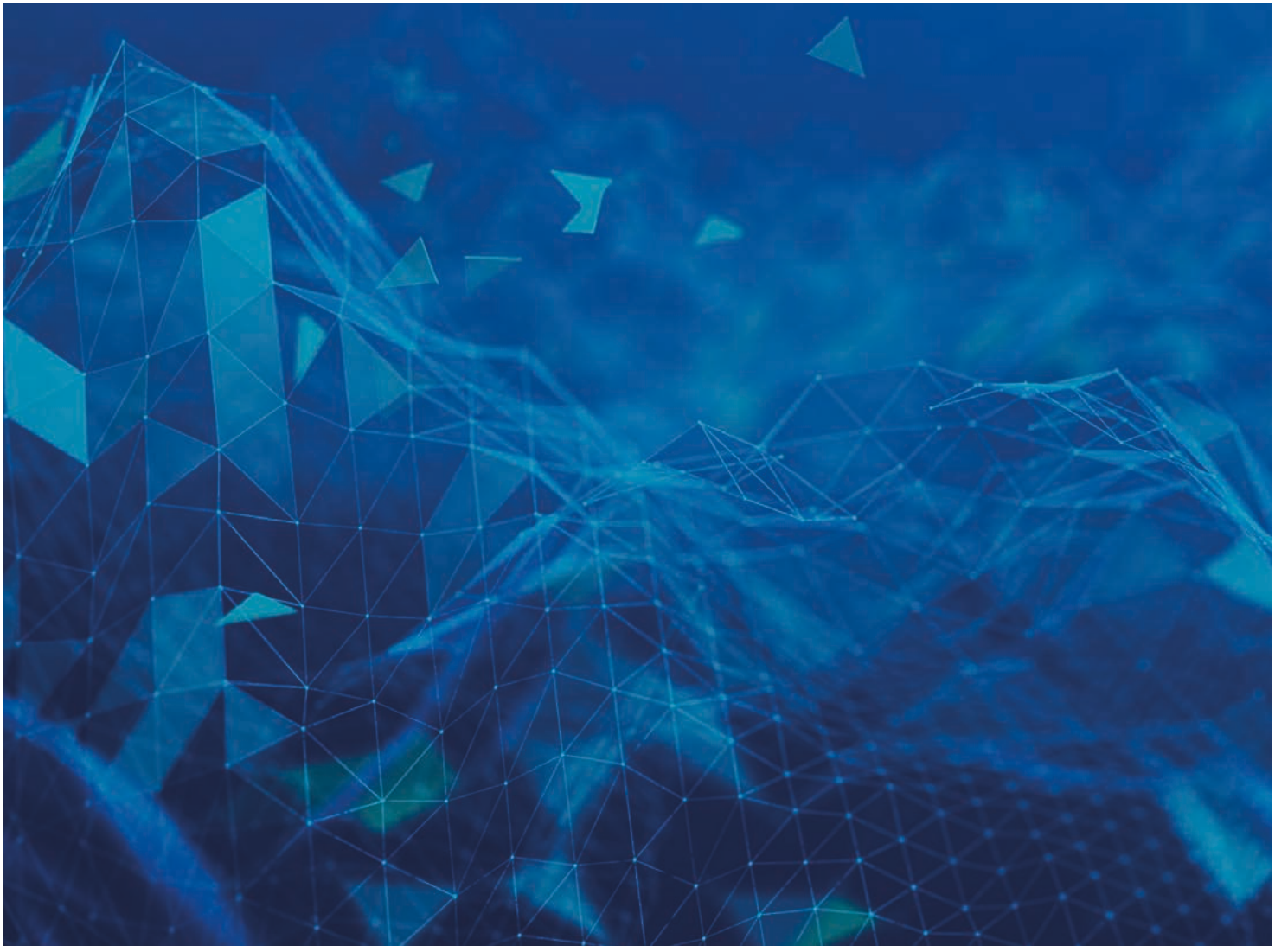
THE STATE OF EMERGING DATA 2025

Tracking a Decade of Data Challenges



EXPERTS WITH IMPACT™

FTI CONSULTING

In 2025, data risks and challenges reached a new all-time high. More general counsel are concerned about the impacts of emerging data sources on compliance, disputes and investigations than at any previous point in time. At the 10-year milestone since emerging data sources began to seep into the corporate risk landscape, this paper examines the top 10 trends taking shape today and impacting digital risk on the horizon.

Nearly all litigation and regulatory investigations are now impacted in some way by the challenges of preserving, extracting, analyzing, reviewing and producing data from cloud suites, chat applications and collaboration platforms. With this, legal professionals are realizing that traditional workflows and case law that were once cornerstones of standard practice can no longer provide default guidelines for digital forensic and discovery methodologies.

The General Counsel Report 2025 found that 31% of chief legal officers rank increased data volumes as one of their top five legal risks. Beyond the issue of growing data stores, 88% of legal department leaders are also specifically apprehensive about the risks surrounding emerging data sources — of those, 59% said they are "very" or "extremely" concerned, compared to 30% in the prior year report.

Data Worries

65% of organizations are **minimally or not at all prepared** to handle issues related to emerging data sources.

47% of legal teams have experienced **new governance, compliance and discovery challenges** associated with cloud, chat and collaboration tools.

57% have faced new compliance challenges related to their expanding data footprint.

Nearly 20% of industry professionals expect emerging data to be the **top focus** of e-discovery this year.

SOURCES: *The General Counsel Report 2025*; eDiscovery Today 2025 State of the Industry Report

An Evolution of Emerging Data Sources

2015
30% of corporate data had migrated to cloud storage or enterprise suites. In-house counsel worried about bring your own device challenges and the implications of big data.

2017
Microsoft Teams launched, joining a growing wave of tools to offer increased collaboration and productivity for enterprise users.

Remote work forced by the pandemic spurred a near-overnight explosion in adoption for cloud and collaboration tools, quickly elevating the visibility and usage of emerging data platforms. In turn, the data footprint expanded rapidly, increasing the volume, variety and dispersion of electronic communications and documents that could come into scope in legal matters.

2020
U.S. Magistrate Judge Katharine H. Parker in the Southern District of New York ordered that hyperlinked documents are not equivalent to traditional attachments in the context of e-discovery. The ruling emphasized that hyperlinks do not automatically constitute part of an email's "family group" and that producing all hyperlinked documents alongside their parent emails would be burdensome and not proportional.

2021
Emerging data platforms took hold as critical sources of evidence in disputes and investigations and accounted for approximately 35% of data FTI Technology processed for client matters.

2022
Roughly 80% of the data FTI Technology processed for client matters was from emerging data sources. The judge in a high profile litigation provided some of the most pointed case law to date regarding linked content.

2024
Linked content became a central topic of debate and speculation in discovery circles, with the industry largely split over whether or how hyperlinked content should be considered modern attachments (28% say no, 27% say yes, the rest say it depends).

2025
Discussions about generative artificial intelligence tools as emerging data sources gained momentum.

THE TOP 10 ISSUES IN EMERGING DATA SOURCES IN 2025

Increasing awareness of the problem has yet to catch up to consistent readiness or clear action. Most organizations, law firms, courts and regulatory agencies continue to grapple with foundational concepts in legal hold, defensible deletion, forensic defensibility, proportionality and more. To improve readiness and drive ongoing maturation in this space, legal teams can take steps to understand and address the issues outlined here.

1. Defining a document.

Emerging data sources, particularly chat messages and linked content, have changed the fundamentals of **what a document is** in discovery. Legal teams are naturally focused on continuing to conduct document review faster and more efficiently. While the industry should always be working toward improved efficiency, speed of review isn't the issue causing problems in most modern cases. Rather, legal teams are experiencing cost overruns, or worse, being hit with sanctions, because they missed something critical in discovery, often as a result of following outdated parameters for scoping documents.

Emerging data sources require a broad and flexible view of what constitutes a document: whether it be a traditional attachment, a shared, cloud-based file with multiple authors and versions, a string of chat messages that contain a mix of pertinent and irrelevant conversation threads, or some other record of communications or decisions.

2. Shared access.

Just as the four corners that define a document have changed, the lines have blurred around the definition of a custodian. When so many files are shared and authored between numerous people, there is now a spectrum of different kinds of custodians. The ability to identify the following details are critical to determining whether or not an individual qualifies as a custodian and what kind of custodian that person may be:

- **Who created the document and for what purpose.**
- **What permissions (editor, reviewer, viewer) were provided to additional individuals.**
- **Who set the original permissions, and if they were changed, who made the changes and when.**
- **Who contributed or changed the document and what those actions included.**
- **Who viewed the document or shared it with others.**
- **What communications platforms were used to share the document.**
- **Whether the document was shared as a hyperlink or static attachment.**
- **Whether new versions were created with different permissions and access.**
- **Key points in time related to changes in the content, permissions or sharing.**

3. Chat messaging.

Chat tools and ephemeral messaging applications present countless challenges in conducting investigations and discovery. Short-form messages from apps and collaboration channels are very different from email messages and cannot be threaded in the same way to provide meaningful context about a case. The asynchronous nature of chat messaging means chats can span platforms and lengthy periods of time, jumping between potentially dozens of topics from message to message. In addition to a shorter, more casual communication style, reactions, emojis and images are often interspersed throughout a conversation, adding a layer of subjective meaning and understanding to the dialogue.

Like many avenues of emerging data sources, clear parameters have not been established for how to group conversations, what formats should be used for viewing a chat or channel message outside of its original platform, and how investigative teams can determine the meaning different individuals intended when using reactions, emojis, short-hand and slang in their business communications. Additionally, exported, chat messages, emojis and reactions often render in complex formats that are not searchable or that do not capture all the details or metadata that may be required to fully analyze the meanings.

When chat messages are in scope in a matter, they typically require a current, customized and cautious approach to piece the relevant information together without losing important details or requiring excessive manual work.

4. Linked content.

Last year, a magistrate judge in a high profile litigation provided some of the most clarifying case law to date regarding linked content, indicating that while hyperlinked items are not the same as traditional static attachments, they should be considered attachments in the context of e-discovery production requirements when the functionality exists within the platform to provide this information. Production of metadata for the linked documents was also required in the case, as well as documents linked in certain chat messages. The court recognized technological feasibility limitations related to producing contemporaneous versions of hyperlinked files depending upon where they were stored and encouraged exploration of feasibility to expand the scope for future productions.

While this provides guidance of the role linked content plays in discovery obligations, the nuances matter. Until there is a more substantive body of guidance surrounding linked files, legal teams should not treat or refer to them as the same as traditional attachments.

Additionally, organizations can take proactive steps to reduce risk and the potential downstream burden of handling linked content in discovery. Mapping linked content systems across the enterprise and revising standard e-discovery playbooks to establish workflows and ESI protocols specifically for cloud data sources is a starting point. Additionally, organizations should work with expert providers that have the ability to proactively locate and scan linked content sources to enrich the legal team's understanding of their data and integrate potentially relevant linked materials into e-discovery platforms.

5. Versions.

The proliferation of numerous and persistent versions of a single original file in dynamic document environments with shared access and variable permissions has introduced another layer of complexity for establishing a clear and accurate historical view of content. Versions may reflect countless changes that cannot be collected, analyzed or reviewed through standard linear methods. Moreover, with versions of dynamic, cloud-based documents, come nuanced shared access roles and permissions. This adds complexity to custodian identification and data authentication, while also increasing data volumes.

> *Cloud-based files lack many of the readily identifiable characteristics that often make authentication possible. Specifically, files uploaded to remote servers are not necessarily shared with other users, which forecloses the opportunity for a recipient to authenticate them. And cloud storage providers may not require detailed profiles of their users, which eliminates another avenue to corroborate ownership of the account's contents."*
>
> — **TIM ANDERSON**, *SENIOR MANAGING DIRECTOR, FTI TECHNOLOGY*

6. AI as a data source.

Up to this point, there have been four pillars of the paradigm shift emerging data sources have driven in legal and compliance: shared access, chat messaging, linked content and versions. **AI should now be viewed as a fifth pillar.** As generative AI tools become more embedded across enterprises, the related data artifacts they create, including prompt-response logs, file access records and AI-generated outputs will inevitably enter the scope of e-discovery for legal relevance.

Currently no case law defines the relevance of AI-generated content as evidence in disputes and investigations, which will create legal ambiguity and an array of practical challenges with preserving, collecting, analyzing and reviewing these new data artifacts when they do come into question.

7. AI as a solution.

While AI will create new challenges as an emerging data source, AI tools will also present solutions to help improve data enrichment and provide new workflows for handling linked content and chat messages. For example, when applied alongside expert-led workflows and solutions, generative AI can help to quickly and defensibly enhance contextualization and summarization of chat threads to identify central themes.

These tools can analyze and interpret rich media including images, GIFs, reactions, emojis, audio clips and other modern communication formats across a wide range of platforms. Subject-matter experts and digital forensics investigators can leverage generative AI to enrich metadata, extract key facts and uncover relevant patterns. These capabilities support efficient data processing in a closed, secure environment to mitigate risk, reduce data volumes and accelerate access to critical insights.

8. Continuous evolution and emerging standards.

Emerging data sources require continuous improvement, education and innovation, as they change daily. Software providers update their platforms all the time. Startups launch new productivity technology that agile teams readily adopt without recognizing the potential legal and compliance implications. Organizations build proprietary tools to improve their employees' output. IT departments reconfigure settings to meet specific needs within the IT architecture. Against this backdrop, it's difficult to set consistent discovery standards the legal field can agree upon and adopt.

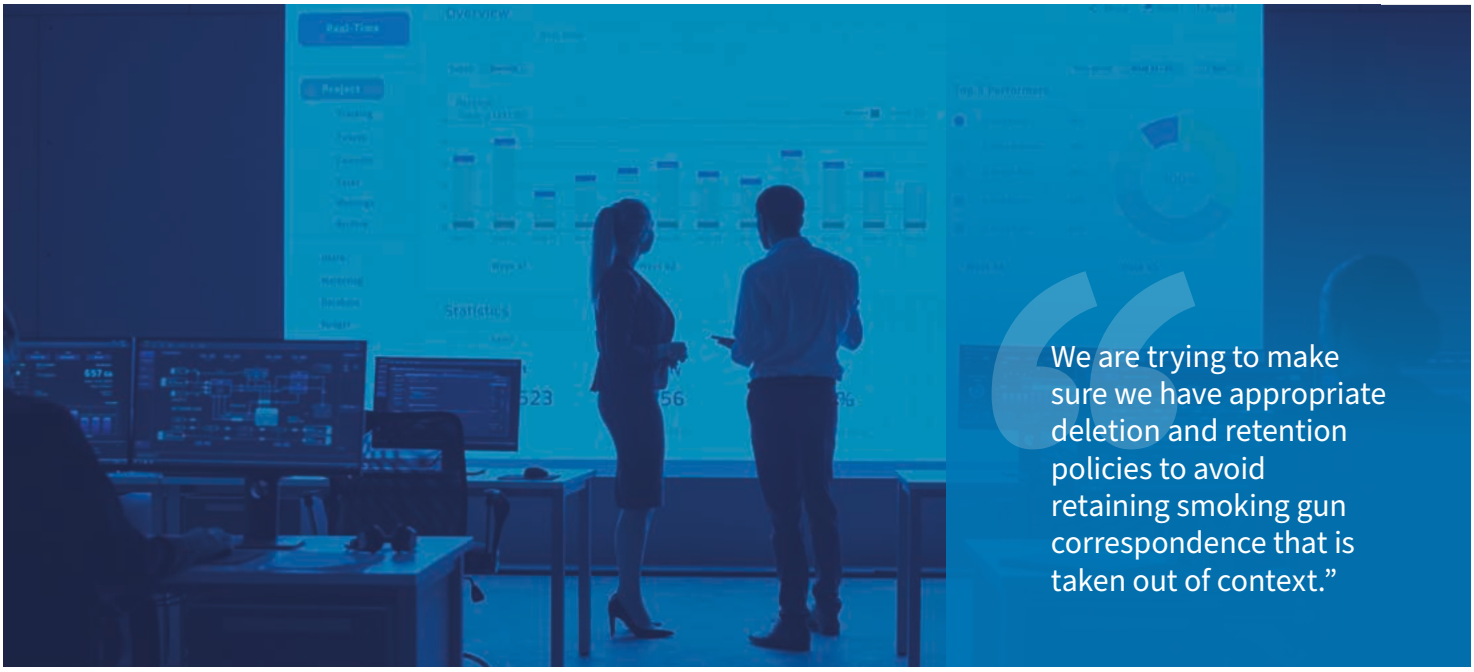
FTI Technology's Emerging Data Sources team solves this by collating intelligence and learnings from experts all around the globe who are interacting with a wide range of data sources across a variety of legal and regulatory matters every day. These experts convey their findings to their worldwide colleagues, work together to rapidly develop solutions and create a feedback loop that supports real-time knowledge enrichment of the latest emerging data sources issues. To the extent possible, legal teams can aim to replicate a similar network of experts in the field, working to increase communication and collaboration across their organization, with outside counsel and with trusted expert advisors.

For industry standards, The Sedona Conference has issued draft guidance for collaboration platforms and some case law has been established. This is a good start, but even incremental progress is not likely to keep up with the pace of change. Therefore, legal teams need to be proactive in anticipating the issues that may arise and doing what they can to maintain a readiness position.

9. Cloud-first forensics.

When collecting from cloud sources, legal counsel and investigators who must maintain established standards of forensic defensibility need a new set of methodologies and workflows to ensure that data integrity — including preservation of metadata and maintaining chain of custody — is maintained throughout the collection process and in a manner that can withstand legal scrutiny. To maximize the admissibility of electronic evidence in court and reliability for investigative purposes, legal teams should ensure their cloud forensics methodologies and tools include capabilities such as:

- Integration with leading cloud platforms to ensure comprehensive data access.
- Advanced data source discovery using sophisticated software to bring order to varied cloud source endpoints and API call designs.
- Secure authentication protocols and encryption to reinforce data security, user privacy and protection against unauthorized access or corruption throughout investigative and discovery processes.
- Standardized data models to simplify complex data sets from varied sources, normalize common metadata and enhance analytical processes and reporting.
- Audit and tracking, which are vital for maintaining data integrity and forensic soundness.
- Preservation of metadata and support of defensible processes to uphold admissibility in legal proceedings and withstand challenges related to data authenticity and integrity.



> We are trying to make sure we have appropriate deletion and retention policies to avoid retaining smoking gun correspondence that is taken out of context."

10. Modern information governance and litigation readiness.

Most organizations have foundational data retention and deletion programs and legal hold playbooks. However, these are typically outdated and have not been designed to fit with the realities of today's data landscape, including cloud environments, chat messages, linked content, document versioning, access controls and AI-generated data artifacts within the enterprise. Areas that should be addressed to create a defensible retention and deletion program oriented to emerging data sources include:

- **Data mapping**
- **Inventory of legal holds to identify and consistently preserve data responsive to a hold, no matter where it resides. Data within chat, cloud and collaboration platforms that should be on hold may be inadvertently deleted without proper controls.**
- **Revised scoping of preservation requirements including custodian and data source identification, retention schedules, compliance monitoring and documentation.**
- **Defensible disposal policies and procedures that are inclusive of all cloud, chat, collaboration and emerging data sources.**
- **Monitoring for changes in compliance tools and archives, to identify impactful changes to data elements and field structures to avoid downstream e-discovery problems.**

SNAPSHOT OF CHANGES IN MAJOR APPLICATIONS

Microsoft 365
400 updates just to Copilot in 2024; dozens of additional updates in 2025, including new UI for Purview eDiscovery, linked content inclusion, version labeling and new legal hold options.

Slack
More than 30 updates reported in the second quarter of 2025; recent changes impacted legal hold, links to attachments, channel export, metadata fields and custom retention settings.

Google Workspace
Hundreds of updates in 2025 to date, including retention rules based on drive labels, client-side encryption and AI summarization availability in Google Vault.

IQ.AI BY FTI TECHNOLOGY FOR EMERGING DATA SOURCES

When applied in tandem with FTI Technology's Universal Messaging Platform, IQ.AI can group content from short-form messages thematically, regardless of the system it originated from, the message flow or format, enabling experts to piece together what happened, when and who was involved, even across fragmented data sources.

Conclusion

All too often, legal teams jump into discovery assuming their traditional workflows will apply no matter the data sources in scope. When issues arise downstream, they are forced to either backtrack or start over completely. Given the criticality of defensibility in discovery, investigations and information governance, it is imperative that legal teams understand communication and collaboration systems in modern organizations. Mitigating digital risk, ensuring data compliance or conducting a defensible review requires understanding parameters around what's available and possible across emerging data sources, as well as identifying how individuals are using these systems enterprise-wide. That technical knowledge is essential to bringing back a semblance of predictability to the discovery process, and to inform the legal team of the options for handling each unique data source. Moreover, to inform an appropriate balance between satisfying the courts and managing risk tolerance regarding the scope of discovery.

Organizations must be aware of the new realities of emerging data sources in the context of legal and compliance, and develop an understanding of how to avoid common pitfalls that may quickly derail a case.

TIM ANDERSON
Senior Managing Director
+1 (415) 293 4483
tim.anderson@fticonsulting.com

COLLIN MILLER
Managing Director
+1 (915) 801 8997
collin.miller@fticonsulting.com

JERRY LAY
Senior Managing Director
+41 78 668 68 68
jerry.lay@fticonsulting.com

The views expressed herein are those of the author(s) and not necessarily the views of FTI Consulting, Inc., its management, its subsidiaries, its affiliates, or its other professionals. FTI Consulting, Inc., including its subsidiaries and affiliates, is a consulting firm and is not a certified public accounting firm or a law firm.

FTI Consulting is the leading global expert firm for organisations facing crisis and transformation, with more than 8,300 employees in 34 countries and territories. FTI Consulting is dedicated to helping organisations manage change, mitigate risk and resolve disputes: financial, legal, operational, political & regulatory, reputational and transactional. FTI Consulting professionals, located in all major business centres throughout the world, work closely with clients to anticipate, illuminate and overcome complex business challenges and opportunities. © 2025 FTI Consulting, Inc. All rights reserved. **fticonsulting.com**