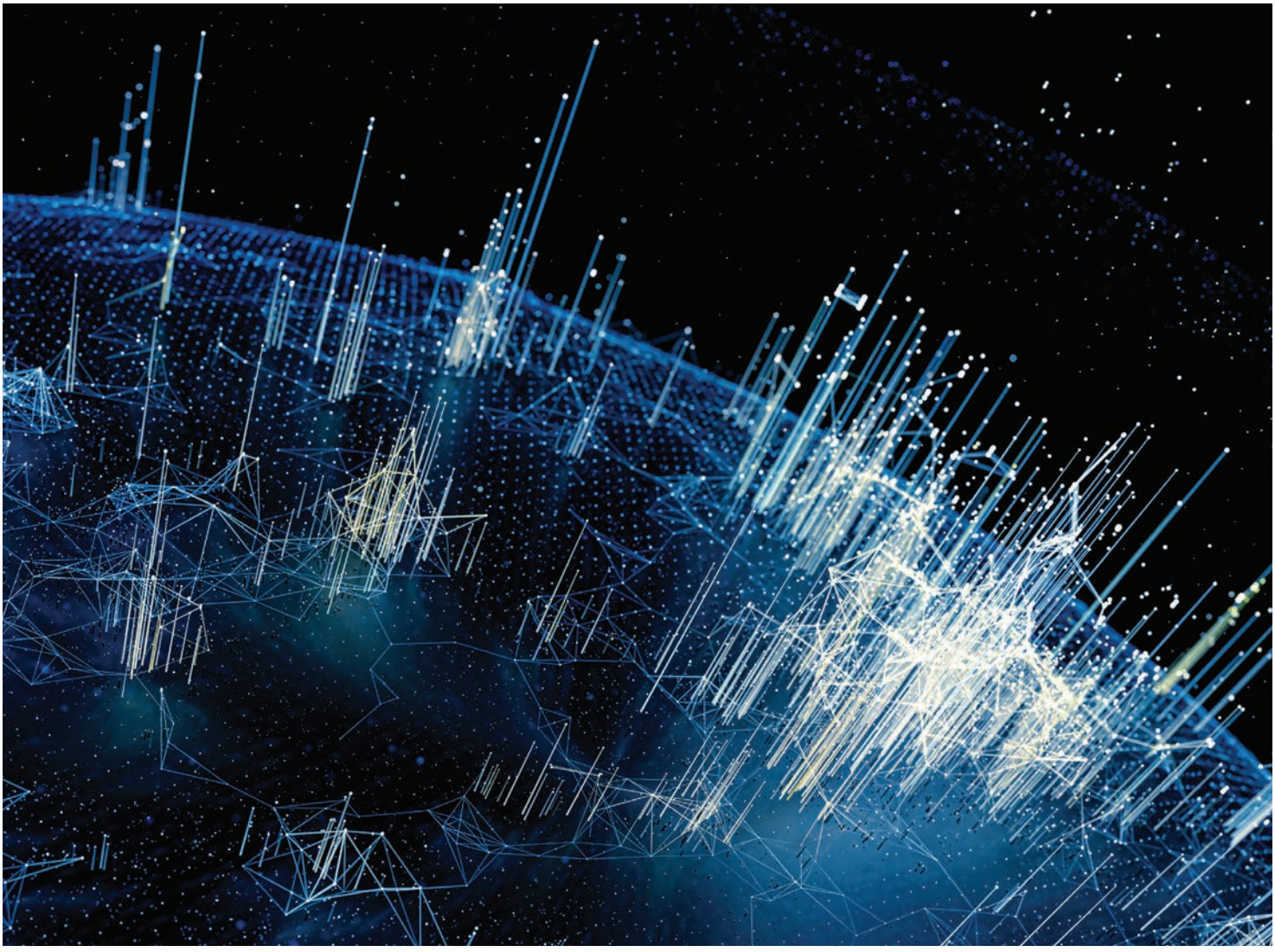
AN FTI TECHNOLOGY REPORT

Navigating Cloud-First Digital Forensics Across Asia Pacific Disputes and Discovery



Cloud data sources and the prevalence of messaging applications used within the workplace have spurred a new cycle of disruption in digital forensics, investigations and information governance. Overall user numbers for cloud-based platforms, and the variety of applications tapped for business purposes, will only continue to grow, especially as tech giants innovate with artificial intelligence and release new collaborative interfaces. The upshot is that counsel will grapple with how to uphold accepted principles for forensic defensibility, electronic evidence collection, data preservation and information governance across vast, complex, rapidly changing populations of data.

This paper will cover the fundamentals of what emerging data sources are, the specific legal and technical challenges around them, solutions for handling them in a forensically defensible way and recent court decisions that will help guide legal teams as they adjust to this new paradigm.

Defining Emerging Data Sources

An emerging data source is any cloud-based platform, collaboration tool or messaging application used for business purposes and communications. Across Asia Pacific, the most known are Microsoft 365, Google Workspace, Box, Dropbox, Slack, Zoom and WhatsApp, WeChat, DingTalk and Lark. While most people are familiar with these tools and use them with ease in day-to-day work and personal communications, their backend complexities and technical nuances have created several persistent challenges in the context of disputes, discovery, regulatory compliance and information governance. Moreover, data and document volumes are growing all the time, and the variety of emerging data sources is expanding all the time via new platforms, and also in the numerous formats that exist within each platform and the types of metadata they include.

As a result, specific challenges have arisen, including:

Access and preservation

Emerging data sources are prevalent across company systems, employee computers and personal devices being used for work. Gaining access to and preserving traditional company documents and email is a routine process, however doing so for encrypted WhatsApp data on an employee's personal phone is not. Aside from the many legal and practical challenges of accessing certain applications and devices, traditional forensic tools and workflows do not have plug-and-play capability to execute a complete, forensically-sound collection for many cloud sources.

Duplicates, near-duplicates and changeable files

With dynamic documents (i.e., linked content, or sometimes referred to as modern attachments), there are a vast number of systems creating and saving numerous versions of a single document. As legal teams are tasked with collecting and reviewing documents linked in messages, they must determine and record which version of that document is relevant, whether there's a need to review and/or produce all versions, and if changes from version-to-version are relevant or need to be analyzed against a broader dataset.

Metadata extraction

With traditional data sources such as email and documents, metadata (i.e., email from, to, cc, sent date, subject, document's created date, last access date and last modified date) can be easily extracted via forensic tools and e-discovery tools automatically. For emerging data sources, metadata extraction is difficult and requires deep understanding of the data structure and the technic of scripting to parse out metadata information.

Authentication

Traditional processes for authenticating information come into question when cloud-based sources are involved. For example, when data is connected from numerous, separate repositories into a single chat message, how is the original source determined and verified? How can teams document and defensibly explain the flow of data coming in and out of a record stored in a collaboration application? Authenticating to cloud endpoints is another challenge. Each cloud endpoint implements a variety of authentication protocols, and while standards have emerged for authentication and authorization, cloud providers implement a variety of these options in varying degrees.

Custodian identification

The process of appropriately or efficiently targeting individuals/custodians and their activities must now also be gauged through the lens of their access, permissions and participation in certain documents or chat channels. Can a single custodian be defined when a single document contains content from multiple contributors in a shared workspace? Does membership alone, without active participation in a communication channel constitute custodianship? Generally, even the notion of a custodian is problematic when dealing with cloud sources. Usually, cloud sources have entities, like users or shared drives. Attempting to identify a custodian obscures the source and often doesn't work for modern cloud platforms, which runs contrary to traditional forensic collection philosophies.

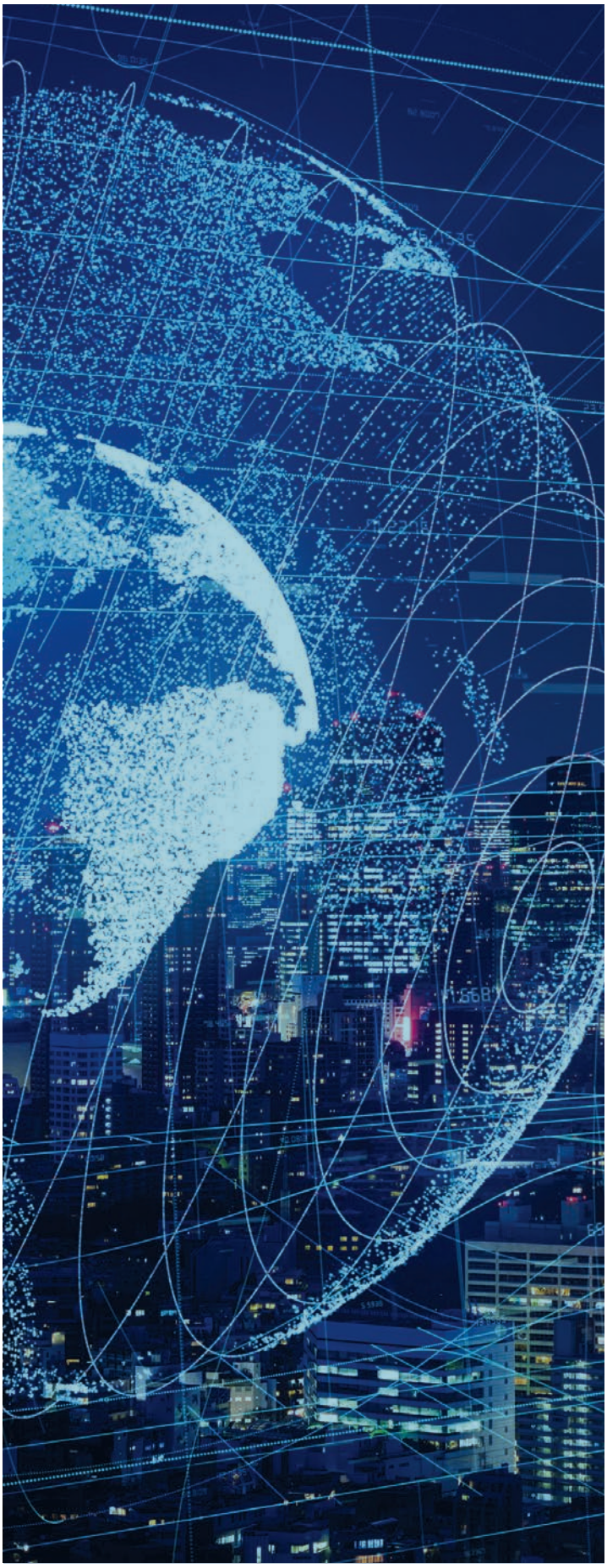
Data Retention and Defensible Disposal

For many of the same reasons that forensic investigation workflows are difficult across emerging data sources, information governance is also challenging. When data can't be easily collected and preserved, it's also difficult to enable automated and sustained defensible retention and disposal processes. Yet without those, volumes and data risk continue to grow, creating a reinforcing cycle of data bloat and complexity.

Achieving Forensic Defensibility

Fundamentally, traditional concepts and approaches are no longer applicable to the new digital landscape. Still, it is possible to remain faithful to digital forensic principles of admissibility, authenticity, repeatability and chain of custody, while leaning into new approaches and tools that are better aligned to the technical nuances of emerging data. For example, with the right approach, legal teams can still authenticate a point in time based upon metadata within a linked document, even if a static, local attachment isn't available.

APIs (application programming interfaces) are an important element in the quest to modernize approaches to defensibility in forensic investigations and e-discovery. Millions of developers and billions of users globally rely on APIs provided by credible technology companies to provide secure, consistent and trustworthy methods for extracting and exchanging data and developing solutions. They are foundational infrastructure in the modern data ecosystem. Understanding how they function and documenting the protocols in place and the details about the API when using them, can provide a strong foundation for defensibility.



Particularly as many emerging data sources do not have export features or have significant limitations in the format and throughput of data exports, APIs are integral to collecting data in an investigation and converting it into a format that can be ingested and reviewed. In a forensic collection, APIs may be used for example, to retrieve a native file and other content, capture specific metadata and/or retrieve versioning information.

Maintaining Data and Artifact Integrity in Investigations

The heart of forensic defensibility in investigations lies in maintaining data or artifact integrity. During any investigation, the original data collected must remain unaltered to maintain its potential evidentiary value. This integrity is particularly important when dealing with cloud-based platforms, as they often contain unique or supplemental artifacts that might not be present in traditional storage media.

Artifacts unique to cloud platforms can range from user access logs, network traffic data, application metadata to remnants of shared access environments, versions of documents and chat messaging data. These artifacts are crucial components of investigations, as they can provide insights into user behavior, data modification history and other key elements. In many cloud platforms, it is common for activity logs or version history to exist for a trailing time period. Without swift action to preserve these logs in their entirety, data critical to an investigation may be destroyed permanently.

By adopting forensically sound data collection practices and focusing on data integrity, digital forensic specialists can ensure that the collected information retains its full evidentiary value, making it reliable and admissible in legal proceedings. This level of diligence also offers an added layer of confidence across the broad scope of digital forensics assignments, as it reassures stakeholders and legal teams that the evidence collected is both authentic and untampered.

Conclusion

The ever-changing nature of today's digital landscape, marked by the exponential growth of cloud data sources, collaboration tools and messaging applications, has fundamentally reshaped the field of digital forensics and other adjacent workstreams.

Enhancing digital forensics and across cloud data sources is both an imperative and a complex task. Meeting the challenges posed by emerging data requires a symbiotic relationship between understanding the nature of the data and the integration of new technologies and approaches that respect the principles of forensic defensibility.

SANDEEP JADAV
Senior Managing Director
+852 3768 4730
sandeep.jadav@fticonsulting.com

CHELSEA YE
Senior Director
+86 21 3892 6129
chelsea.ye@fticonsulting.com