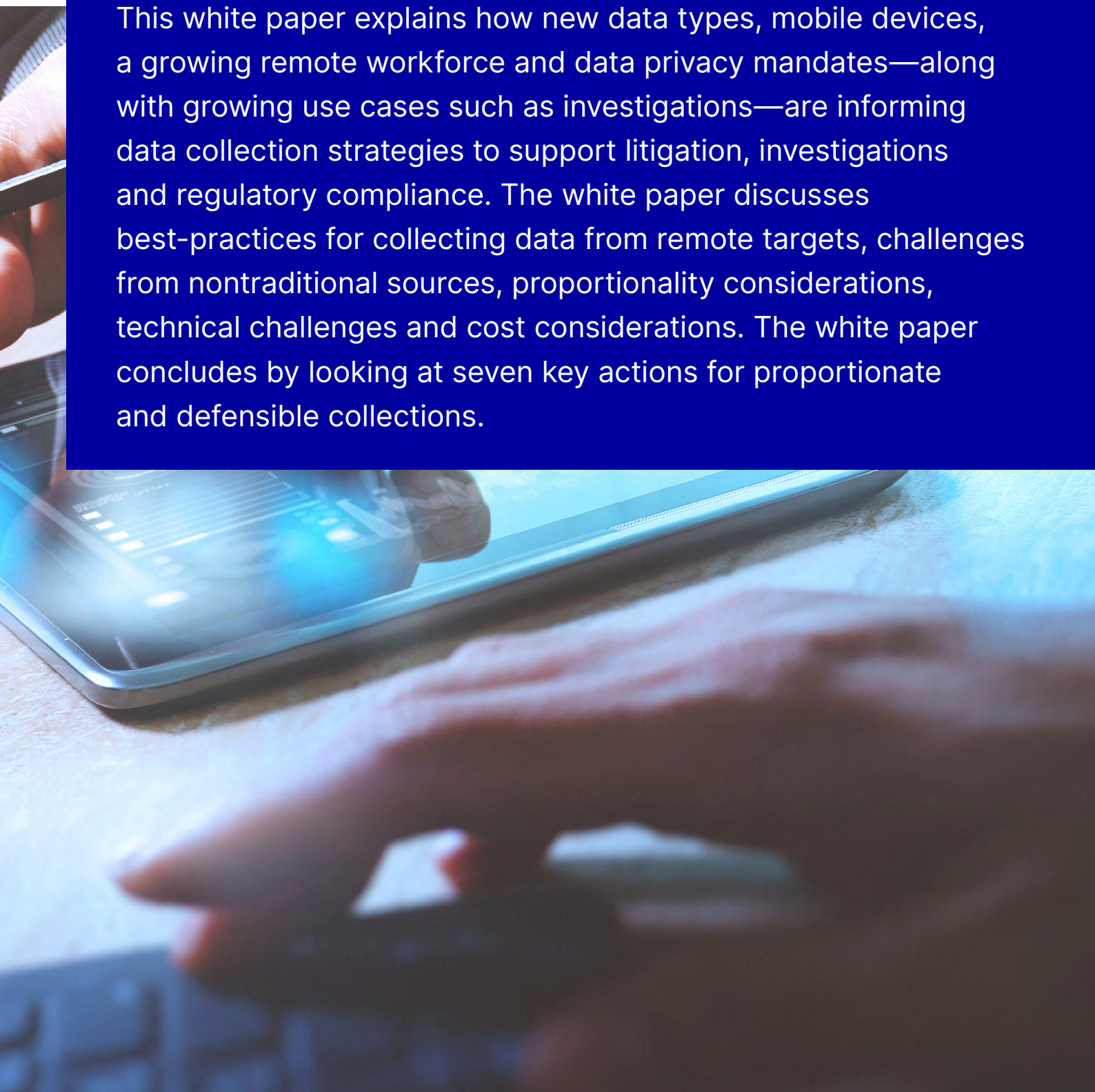
opentext™

WHITE PAPER

Modern ESI Collections: New Imperatives and Best Practices

This white paper explains how new data types, mobile devices, a growing remote workforce and data privacy mandates—along with growing use cases such as investigations—are informing data collection strategies to support litigation, investigations and regulatory compliance. The white paper discusses best-practices for collecting data from remote targets, challenges from nontraditional sources, proportionality considerations, technical challenges and cost considerations. The white paper concludes by looking at seven key actions for proportionate and defensible collections.



opentext™

Contents

opentext™

Executive summary

Litigation, investigations and regulatory response requirements for collecting electronically stored information (ESI) that might potentially contain evidence has evolved dramatically since the introduction of electronic discovery. Data continues to expand both in volume and scope and comes in all different types and formats. Data is no longer exclusively found on local computers or in-house servers; instead, it's in the cloud or on mobile devices like tablets or smartphones that may not even belong to the organisation. This, in turn, raises new concerns about how to balance an organisation's business needs with individual data privacy rights. Furthermore, as the workforce becomes increasingly remote, Slack™, Microsoft® Teams®, Zoom™ and other chat and communications platforms have proliferated, often outside the control of corporate IT.

Organisations need effective and legally defensible ESI collection methods to support legal matters. Given ever-expanding data types and volumes, organisations also need to consider how to narrow the sheer volume of data into a manageable collection size that is proportionate, cost-effective and expedient.

This white paper discusses the changing nature of business data, new challenges and best-practice strategies for ESI collections, including an increasing focus on remote collections. It concludes with seven key steps for proportionate and defensible collections.

Introduction

Data has changed since the advent of eDiscovery and electronic disclosure. Data no longer looks the same, nor does it reside exclusively in discrete, well-defined local repositories. Organisations need ESI collection methods in support of litigation, investigations and regulatory compliance matters that are effective and legally defensible. But with new data sources such as chat, and skyrocketing data volumes, teams also need collection strategies for acquiring data expansively to broadly preserve data, while winnowing down data collections to constrain the effort and cost of reviewing the data. Under the gun with rising legal matters, organisations need the ability to gain rapid insights into their many sources of data, whatever their origin.

The prevalence of working remotely, which was already a trend pre-COVID 19, has increased drastically due to the pandemic and is expected to continue for some time. Working remotely intensifies the challenges around each of the issues mentioned above, while shifting a large volume of data from endpoints within corporate firewalls to laptops and desktops in home offices. These home-based computers and laptops are harder to control and are more likely to be owned by employees, necessitating that the organisation balance its business needs with the data privacy rights of individual employees. The use of cloud-based data repositories also has increased, putting further pressures on legal and IT teams seeking to collect comprehensively across data sources.

This white paper provides insight into:

- The changing dynamics that legal and IT teams need to be aware of;
- Key considerations for proportionate and defensible collections focused increasingly on nontraditional sources of ESI and remote people and devices; and,
- Strategies to conduct remote collections cost-effectively and expeditiously.

The changing dynamics of data collection

We are not merely waxing nostalgic when we note that data collection used to be simpler. When electronic disclosure was nascent, most discoverable information could be found in forms that may have been digital but that emulated paper such as emails, word processing documents, and spreadsheets. That information was located entirely within the organisation's walls, on fixed-location computers and in-house network servers. Data collection projects could be performed almost exclusively in the office. While organisations had an obligation to protect data, there were fewer threats to data and fewer legal requirements for its protection.

In those days, data also was rarely encrypted. Today, encryption is standard on many devices through applications such as Microsoft® BitLocker® and Apple® FileVault®. Additionally, messaging apps such as WhatsApp™ often employ end-to-end encryption. This compounds the challenges of collecting data. Not only must the data collection tool be able to connect to the various devices, the data on those devices may also require decryption, either subsequently or through the device itself.

Let's take a brief look at how today's environment has pressed the fast-forward button on change.

Ephemeral data and new data types

Much of the data that needs to be collected today—whether for discovery, disclosure, or, as we'll touch on in a moment, investigations—has no direct counterpart in the world of paper. It's not a traditional "document" that might exist in a word processing program or an email that's essentially a digital version of a posted letter. Today's discoverable data might include messages and integrated notifications from collaboration applications like Microsoft® Teams®, Slack™, videos, outputs from Internet of Things (IoT) devices and countless other new types of ESI. Some of that data is ephemeral, or short-lived: it's rapidly deleted or replaced by new incoming data, such that any preservation or collection effort must be undertaken promptly. Ephemeral data may not exist, or be recoverable in a readable state, from centralised servers. Even where online backups do exist, these are often encrypted and can only be decrypted using a system token that is stored on the endpoint device.

Data in the cloud and an increasingly remote work force

Corporations and government agencies alike have been moving away from stationary computing resources like desktop computers, in-house servers and intranets. Nowadays, most employees do their work on some combination of mobile devices such as laptops, tablets and smartphones. That shift was made possible by both technological advancements in mobile devices and the shift to cloud rather than on-premise storage. Cloud computing had become widely used by 2018, when 96 percent of corporations were using the cloud[1] for at least some of their operations.

Not surprisingly, the coronavirus pandemic has pushed organisations even further into the cloud as the workforce becomes increasingly remote. The Flexera 2020 State of the Cloud Report found that 30 percent of enterprises had "significantly higher" cloud usage due to their transition to working from home, while another 29 percent reported that their usage was "slightly higher" than normal. Nor are these changes likely to disappear when the crisis is over; the success that businesses have had with a remote workforce is expected to give rise to an enduring shift in how and where organisations and their employees get work done.

This gives rise to three problems for ESI collection. First, collection efforts are no longer restricted to one physical location but must instead span both on- and off-network locations. Second, the widespread use of personal devices for business purposes makes it even more difficult for organisations to collect data that belongs to the business without trampling the privacy rights of individual employees. Third, remote employees working from home often have limited bandwidth and do not always keep their systems powered up. This requires that data collections systems are able to adapt to lower throughput rates and sporadic connectivity to complete collections without timing out or having to start the process over when a disruption occurs.

Data privacy regulations and laws

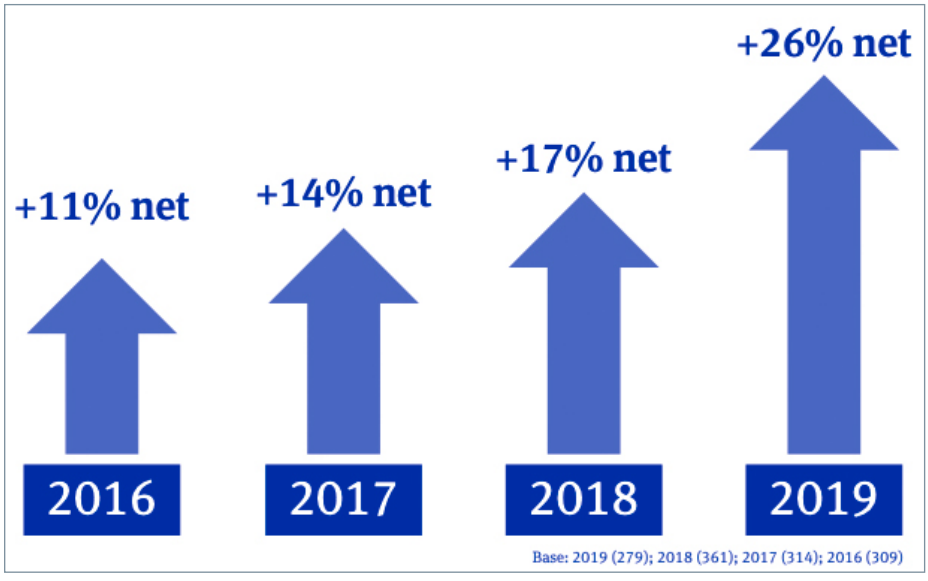
In the past few years, a patchwork of data privacy regulations have proliferated, from the EU's General Data Protection Regulation (GDPR) and U.S.-based California Consumer Privacy Act (CCPA) to the Japanese Act on the Protection of Personal Information (APPI) along with myriad other state and local laws designed to give individuals specific rights regarding their personal information. These regulations have increased the burden on organisations to ensure that their data collection strategy gives due respect to individuals' data privacy rights.

For example, many organisations—especially in the new work-from-home world—either allow their employees to access corporate email accounts, Slack channels and company documents from their personal devices or are unaware that their employees are accessing these systems from their personal computers. While any corporate data on those devices may be discoverable and may therefore need to be defensibly preserved and collected, employees' privacy rights therefore will need to be addressed when collection methodologies are decided upon. Collection practices must therefore take a holistic approach, balancing the legal and business needs of the organisation with the data privacy rights of the individual device owner.

In short, data has become simultaneously more complex and more widespread, leading to new privacy considerations. At the same time, organisations are seeking to collect data for more than just litigation, adding another layer of pressure on data collection approaches.

New—and more—litigation on the horizon

Litigation has been trending upward—accompanied, of course, by electronic disclosure—for several years. According to Norton Rose Fulbright's 2019 Litigation Trends Annual Survey, there was a "sharp rise in the proportion of organisations predicting an increase in disputes" for 2020. Thirty-five percent of responding organisations expected an increase, while just 9 percent predicted a decrease, for a net increase of 26 percent—considerably higher than the 17 percent differential in 2018 or the 11 percent difference in 2016. The survey concluded, "The world is in a period of uncertainty where the full effects of trade wars and economic cycles remain unknown. Uncertainty breeds fear and we are seeing the results of that fear in organisation's predictions for increasing dispute activity."



Source: Norton Rose Fulbright 2019 Litigation Trends Annual Survey

Those predictions were made before COVID-19 emerged as a global threat, sending uncertainty and fear through the roof. In addition, it's expected that legal issues and claims surrounding the pandemic and accompanying economic downturn will lead to a new surge of cases involving employment, workers' compensation, contract disputes and other litigation issues.

1 CIO.com "IT governance critical as cloud adoption soars to 96 percent in 2018"

Rise in investigations

But litigation and regulatory response aren't the only use cases for the collection of ESI. Just as litigants seek to get a handle on relevant information to assess the strengths and risks of their case, organisations also want to understand what their data indicates about internal behavior to identify risk—especially in light of a heightened regulatory framework—before it turns into a fully blown litigation matter, to remediate as quickly as possible or take other necessary courses of action.

In fact, ESI collections to support investigations has become a tremendous growth area. Investigations pose new challenges; for example, timelines tend to be truncated, requiring quick insights to inform rapid decisions—and consequences, such as in the case of IP theft, can be dire to the company if not detected quickly. Further, overburdened internal corporate IT staff may be slow to support a collections effort, and significantly, custodians may be less likely to provide timely and complete information about data related to investigations if they think it may negatively impact themself or a colleague.

Investigations that may trigger a need for ESI collection include internal investigations, compliance and regulatory investigations. These can encompass issues such as HR disputes involving furlough, discrimination or harassment, due diligence investigations in the course of mergers and acquisitions and pre-assessment of litigation claims. The rise in cybersecurity threats and IP theft are another significant growth area for ESI collections. Collecting the data involved is essential to understanding the volume and sensitivity of the data to assess the scope and implications.

For a host of reasons organisations need to have the ability to rapidly survey their data and collect a wide variety of information, from a broad array of sources, that may be relevant to decisions that the organisation must make regarding litigation, regulatory compliance and internal dealings.

Key considerations for a proportionate and defensible remote collection strategy

Non-enterprise IT systems

Collections challenges are substantially compounded by working remotely. These include:

- IT policies put in place in office settings may be overly or not sufficiently restrictive. This can lead to either barriers for remote workers or risks to corporate assets and the data on them as employees seek to circumvent strong restrictions;
- Data archiving enforced behind the firewall can fall off substantially because policies are not as easy to enforce on employees' home computers;
- Getting access to the right IT staff has become harder for remote workers and project managers that rely on IT to help execute collections alike;
- Employees may have slow Internet connections that can slow down collections processes;
- Remote devices can be powered down or disconnected without notice, putting a pause on collections processes and potentially causing collections to fail mid-process;
- Furloughed employees can't be asked to assist because even powering up a laptop is a work activity; and
- The use of corporate virtual private networks (VPNs) that allow employees to access company systems can create a bottleneck in data transfer speeds, assuming that the custodian can get a connection—many organisations don't have suitable VPN capacity for 100% of their workforce to be connected simultaneously.

Avoid the cost of repeat collection

Effective data collection is a key means of satisfying the duty to preserve data when litigation is reasonably anticipated. The critical consideration for cost-effective collection is to ensure that it is sufficiently expansive to encompass any documents that may be necessary to answer the specific questions underlying the case. In addition to making the collection effort more efficient (so there will not be any need to spin up resources more than once), an expansive collection will in fact make the review process more efficient, since it will be more likely to provide the documents necessary to get the full fact picture, rather than leaving evidentiary holes that complicate analysis.

Technology-assisted review (TAR) is designed to quickly home in on relevant data by prioritizing the data most likely to be relevant. As such, the efficiency benefits of collecting expansively far outweigh the effort and risk of trying to collect only perfectly relevant data. Collecting expansively also increases the likelihood that the relevant data has already been collected if case strategy evolves or new custodians are added.

Begin review earlier

Early Case Assessment (ECA) seeks to find and understand the facts that will help counsel understand liability and risks and decide whether to settle or proceed to litigation. ECA becomes possible once data is collected, culled and processed so the more streamlined these operations are, the earlier the efforts to derive insights can begin.

Defensibility considerations

Collections should also include a comprehensive audit chain by using physical or logical forensic image containers for data sets across the collection and subsets of the data. Audit chains use cryptographic hashes that generate a unique "fingerprint" with which to identify the data and show whether the data has been modified anywhere within the collection process. Audit chains also allow anyone to follow what has happened to data from the point of collection, allow the processed data to be verified against what was supposed to be collected, and provide a guide to any issues that need to be corrected. Audit chains go hand-in-hand with scoping documentation and written decisions on the data intended to allow for review of what was available versus what was collected.

The use of industry standard tools and techniques, avoiding any alteration to metadata, and maintaining chain of custody are other key elements of defensible data collection.

Leverage collection experts as needed

Historically, in-house legal teams relied on the IT department to apply search strings to locate emails for review. Not only have studies shown that Boolean search is not an effective way to locate pertinent documents, but the possibility of missing critical acronyms, product codes and formulaic designations is obvious.

There are further difficulties with putting the collection burden on IT staff, such as missing content in attachments or material which has not been correctly indexed, as well as likely inexperience with deduplication and applying multiple search terms. Further, there is the risk of overlooking an entire category of electronic documents in, for example, a network share or collaboration platform. Finally, many IT staff do not have the necessary skills or software to apply OCR to image-based text, or apply text indexing processes to attachments and containers such as .zip files, and other steps to comprehensively search and collect documents, as these capabilities lie outside core competencies expected of IT personnel.

To avoid these potential impediments to a thorough collection effort, the scope should be designed expansively. Modern data collection not only contains costs through aggregating the effort with experts able to apply a broader range of tools, collection experts are also more experienced in collecting newer forms of data. Ephemeral data is a primary example in which collection techniques must be persistent to capture this short-lived data but unobtrusive to avoid hi-jacking local computing resources. Collecting data from multiple cloud repositories, each with their own criteria, is another example where collection experts can apply their specialised knowledge to contain costs.

It is important to note that collection experts must partner effectively with the organisation's that they serve. There are key decisions such as defining the scope of collections and determining where to draw the line to balance the organisation's needs with the privacy of its employees that collection experts cannot make on behalf of the clients. But collection experts can play a critical role in raising awareness of where key decisions need to be made to help organisations avoid unseen pitfalls.

Use the most appropriate tools for the collection need

Collection experts can also quickly align the right tools to the various requirements. This includes deploying forensic collection tools to collect data expansively without infringing on personal data and adapting connection mechanisms when data is not accessible via corporate VPNs.

Unlike standard IT software utilities like robocopy and rsync, these dedicated forensic tools are designed from the ground up to perform collections in a defensible manner, without making any changes to the document content or metadata during the collection. These tools are subject to extensive testing and peer-review within the industry, and in many cases have been used and approved during criminal and civil cases in numerous jurisdictions.

The output from these tools should align to industry recognised container formats that are read-only so the collected documents can not be modified post-collection. The output should also contain cryptographic verification hash values so that any other party can subsequently receive and verify a copy of the forensic container and be confident that they have an exact copy of what was collected. The forensic images may also capture additional information about the system the document was stored on, which can lead to further avenues of investigation, such as the recovery of deleted documents or previous versions of a document.

It is also not uncommon for a collections expert to conduct a collection of the same system using multiple tools. This process is known as dual-tool verification and is a method of further verifying that the captured content is complete and correct.

Finally, forensic tools typically have extensive logging and auditing capabilities built-in to the data acquisition workflow, meaning errors and warnings will be captured and written into the acquisition log. The practitioner can then work through these to ensure that a full document collection is captured.

7 key actions for every ESI collection

Here are key actions organisations need to take regardless of the ESI collection effort, scope or whether it is remote:

1. **Identify they key custodians and sources:** Collecting proportionally and defensibly relies on knowing who the custodians are related to a matter and where their data resides. Missing a key custodian or data source can leave holes in finding critical communications between custodians, a key financial tracking spreadsheet, a CAD diagram that shows a clear deficiency in the design of a product and other gaps that omit critical data essential to a case. These are just select examples of the myriad types of critical data that can be overlooked.
2. **Justify why not to engage collection experts:** Listing how and why an organisation is able to perform efficient and defensible collections is a direct path to assessing practical capabilities. In most cases, particular deficiencies or skill-gaps will be identified to inform decisions on selecting the right experts to assist and provide guidance to those experts on where help is needed most. Engaging with independent data collection experts also adds an extra level of trust in the methods and outcome.
3. **Create a data map of all potentially relevant sources:** Creating a map of all custodians and their associated data sources helps create a data collection project plan and serves as a reference tool for assessing whether the project is collecting the intended data expansively. The data map is also useful for justifying decisions on what was not collected.
4. **Document the decisions taken:** Keeping record of why particular custodians and specific data was and was not included is a key element of defensibility and streamlines the ability to satisfy questions from opposing council.
5. **Maintain the chain of custody for all collected data:** Data collection is seriously compromised if it results in altered metadata or a broken chain of custody. Both the process and the results must be defensible and forensically sound, with rigorous adherence to the chain of custody. Data collection methods should generate industry standard and legally sanctioned output formats such as EnCase Expert Witness Format E(x)01 and L(x)01 logical images format. Data should also be collected in such a manner that it can be reused across matters.
6. **Use validated tools, processes and experienced practitioners:** Industry standards are well established for both tools and processes. Keeping within the guardrails of industry standards deflects concerns about altered metadata, broken chain of custody, and poorly targeted collections. For added confidence look for practitioners who hold industry recognised qualifications such as the EnCase Certified Examiner (EnCE), and organisations that have been assessed as working to a higher standard through certifications such as the ISO17025 laboratory standard.
7. **Know your (organisation's) limitations:** Conducting proportional and defensible data collection is rife with pitfalls, from being able to access remote devices, understanding the limitations of running searches on live systems using inadequate tooling, through to simply being experienced at mapping common data sources against lists of custodians. A practical recognition of limitations is valuable for selecting and instructing the right collection experts.

opentext™

OpenText Legal Tech

OpenText ESI Collection Services

Conclusion

Given the rising complexity of data types and sources, the surge in mobile devices and remote work and new mandates for data privacy, organisations need to take a more holistic approach to data collection for eDiscovery, regulatory compliance and investigations.

Collections need to be conducted expansively to avoid the effort and cost to re-process collections for missing data. Mapping the collections requirements and creating an inventory of the collected data pre and post collection is key to validate that the correct data has been included. Collecting all relevant data in a single effort is essential to process efficiency while the ability to cull data is critical to containing the cost of data review. Maintaining chain of custody and employing industry standard tools and processes are essential for defensibility.

OpenText offers ESI collection services with EnCE-certified specialists for organisations and their outside counsel. Fast turnaround and highly scalable services, whether remote or on-site—with no interruption to employees means earlier access to the data by review teams. From scoping and physically connecting devices through processing for review and analysis, OpenText will handle the project under the guidance of counsel. Full audit trails and comprehensive chain-of-custody for every project provide complete accountability to defensibly track, store and archive electronically stored evidence.

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