opentext™

WHITE PAPER

Modern Data Collection: New Imperatives and Critical Requirements

Strategies for data collection in today's world

opentext™

Contents

opentext™

This white paper explains how new types of data, mobile devices, remote work and data privacy mandates—along with new use cases such as internal and external investigations—have changed the mandates around data collection. It concludes by outlining the seven critical requirements of modern data collection, which must be comprehensive, fast and efficient, easy to use, discreet, failsafe, defensible and integrated with standard review platforms.

Executive summary

Electronic discovery (eDiscovery)—and consequently, data collection—used to be primarily about email. Now, data comes in all different types and formats. Nor is data exclusively found on local computers or in-house servers; instead, it's in the cloud or on a laptop that may not even belong to the organization, raising new concerns about data privacy. Organizations still need data collection methods that are effective and legally defensible, but they also need to narrow the universe of ever-expanding data into a manageable volume that they can afford to review to gain rapid insights into both litigation and investigations.

This white paper briefly sets out the changing face of business data and reviews the seven key requirements for modern data collection:

1. Comprehensive—collect across all sources of content including endpoints
2. Fast and efficient—collect and cull in a single process
3. Easy to use—crawl in parallel with insight into the data pre- and post-collection
4. Discreet—collect without burden to computing resources
5. Failsafe—automatically adapt when devices drop off of networks
6. Defensible—no alteration of metadata; legally sanctioned output formats
7. Transferable—easy uploads to standard review platforms

Modern data collection demands more than just a quick skim of a hard drive for email. Is your data collection solution up to the task?

opentext™

Introduction

Data has changed since the advent of eDiscovery. It no longer looks the same, nor does it reside exclusively in discrete, well-defined local repositories. Organizations still need data collection methods that are effective and legally defensible. But in the face of skyrocketing data volumes, they also need data collection methods that will immediately winnow down the universe of data into something manageable, not to mention affordable. Under the gun with rising litigation and time-pressured regulatory compliance concerns, they need the ability to gain rapid insights into their many sources of data, whatever their origin.

This white paper explains how new types of data, mobile devices, remote work and data privacy mandates have changed what organizations need from data collection. In addition, it sets out how data collection is no longer just about eDiscovery for active or anticipated litigation—it's increasingly important for investigations, both internal and external. The white paper concludes by outlining the seven critical requirements that modern data collection must satisfy:

- comprehensive, accounting for all data sources;
- fast and efficient, automatically deduplicating and avoiding re-collection;
- easy to use, collecting all data in one process;
- discrete to not disrupt the custodian's work;
- failsafe, correcting for network disruptions and other interruptions;
- defensible, meeting judicial and regulatory standards; and
- integrated, creating outputs that can be easily transferred to standard review platforms.

opentext™

The dynamics of data collection are changing

We're not merely waxing nostalgic when we note that data collection used to be simpler. When eDiscovery was young, most discoverable information could be found in forms that may have been digital but that emulated paper: emails, word-processing documents, spreadsheets and the like. That information was located entirely within the organization's walls, on fixed-location computers and in-house network servers, and work was done almost exclusively in the office. Even when organizations provided laptops, employees rarely took them out of their offices. While organizations had an obligation to protect data, there were fewer threats to data and fewer requirements for its protection.

Many of those characteristics have been shifting gradually, thanks to advances in technology and our more mobile society, but 2020 has pressed the fast-forward button on change.

Ephemeral data and new data types

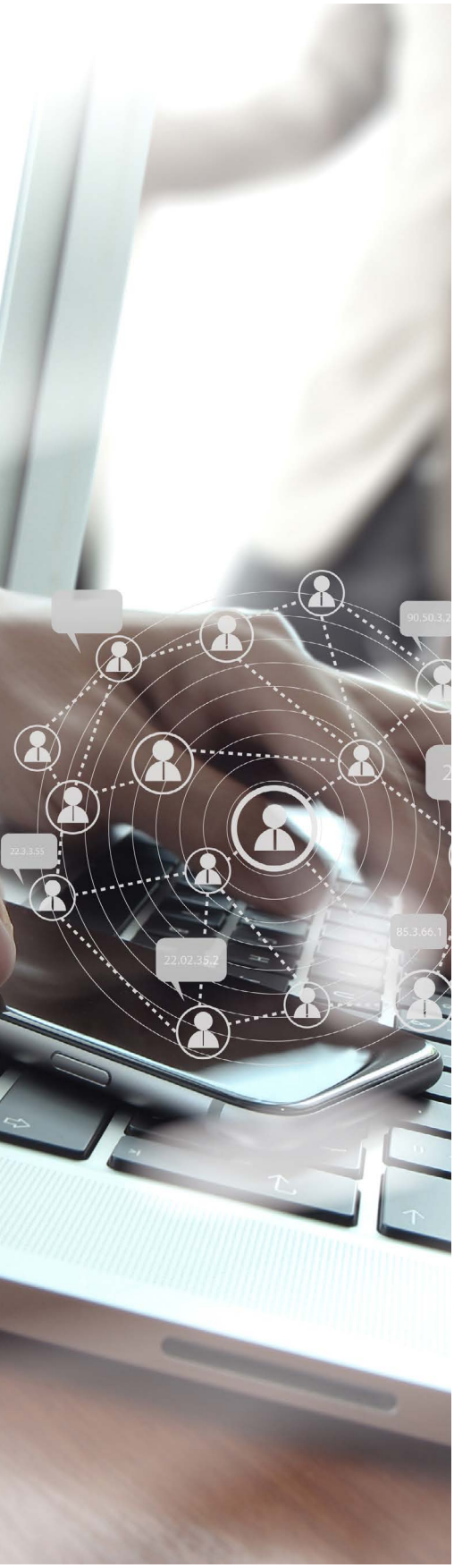
Much of the data that needs to be collected today—whether for discovery or, as we'll touch on in a moment, for investigations—has no direct corollary in the world of paper. It's not a traditional "document" that might exist in a word processing program or an email that's essentially a digital version of a posted letter. Today's discoverable data might include messages and integrated notifications from collaboration applications like Slack, videos, outputs from Internet of Things devices and countless other new types of data. 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. Data collection must now encompass every potentially relevant data type from wherever it may originate.

Mobile data and remote work

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 in-house, storage. Cloud computing had become nearly universal by 2018, when 96 percent of corporations were using the cloud for at least some of their operations.

Not surprisingly, the coronavirus pandemic has pushed organizations even further into the cloud as they've been forced by social distancing mandates to radically increase their remote work capabilities. The Flexera 2020 State of the Cloud Report found that 30 percent of enterprises had a "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 will these changes 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 organizations and their employees get work done.

This gives rise to two problems for data collection: first, collection efforts cannot be 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 organizations to collect data that belongs to the business without trampling the privacy rights of individual employees.



Data privacy regulations and laws

The last few years have seen the proliferation of a patchwork quilt of data privacy regulations, from the EU's General Data Protection Regulation (GDPR) to the California Consumer Privacy Act (CCPA) 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 organizations to ensure that their data collection gives due respect to individuals' data privacy rights.

For example, many companies—especially in the new work-from-home world—allow their employees to access corporate email accounts, Slack channels and company documents from their personal devices. While any corporate data on those devices is discoverable and must be defensibly preserved and collected, the collection methods used cannot infringe on the employees' privacy. Collection practices must therefore take a holistic approach, balancing the legal and business needs of the organization with the data privacy rights of the individual device owner.

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

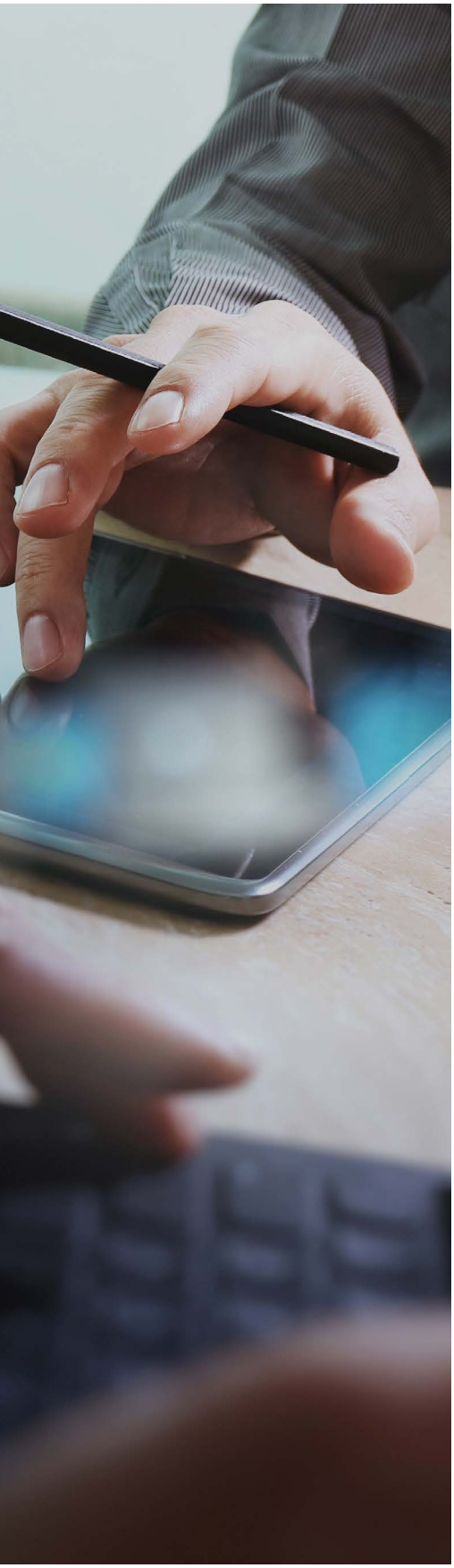
The use cases for modern data collection are expanding rapidly

Litigation was already trending upward—accompanied, of course, by eDiscovery—before the coronavirus pandemic struck. According to Norton Rose Fulbright's 2019 Litigation Trends Annual Survey, there was a "sharp rise in the proportion of organizations predicting an increase in disputes" for 2020. Thirty-five percent of responding organizations 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 organizations' predictions for increasing dispute activity."

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 the unprecedented legal questions 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. Organizations will need to collect enormous volumes of corporate data to resolve these matters.

But litigation-based eDiscovery isn't the only use case for data collection. Just as litigants seek to get a handle on relevant information to assess the strength of their arguments, companies also want to understand what their data indicates about internal behavior, regulatory compliance and potential acquisitions or C-suite hires. In fact, data collection to support investigations has become a tremendous growth area. Investigations pose new challenges. For example, their timelines tend to be truncated, requiring quick insights to inform rapid decisions. Investigations that might trigger a need for data collection include the following:

- government and regulatory agency inquiries;
- internal investigations into reported or suspected misconduct, including HR disputes involving discrimination or harassment;
- due diligence investigations in the course of mergers and acquisitions;
- compliance investigations;



- vetting of new hires or potential C-suite promotions; and
- pre-assessment of litigation claims.

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

The new requirements for modern data collection

Here's what organizations need from their data collection approach.

Comprehensiveness

Modern data collection methods must be comprehensive, consistently collecting data from all relevant sources. This includes physical sources like laptops and desktop computers; cloud sources like Box, Dropbox, Google Drive, Slack and Microsoft Office 365; servers such as Microsoft SharePoint; and all manner of email formats from POP3 and IMAP to Microsoft Exchange and Microsoft PST, whether live or archived. Modern forensic software, such as EnCase™ Information Assurance, include robust connectors to all of these data sources as well as remote agents for collecting data from desktops and laptops to enable comprehensive data collection.

Speed and efficiency

A data collection method must work quickly and efficiently. It should use advanced search filtering to aggressively cull data at the point of collection; use global deduplication to avoid re-collection of data that has already been obtained, perhaps from another custodian; and automatically deNIST data sets to remove machine files and zero-byte files. This minimizes costly, disproportional over-collection in the context of eDiscovery where the cost of review is directly proportional to the volume of data that requires review.

Speedy and efficient data collection also helps organizations identify conclusive information as expediently as possible for investigations where time is of the essence.

Ease of use and insight

There's no reason for data collection solutions to be non-intuitive or cumbersome to use. Modern data collection methods should allow a single unified process to collect data from all sources, crawl multiple target sources in parallel to expedite collection time and enable frequently used criteria to be templated and automated. Additionally, collections should be configurable to target data flexibly in line with requirements. For example, EnCase Information Assurance supports the ability to tailor collections to specific folders anywhere within multi-tiered folder structures. If the objective is to collect all sales contracts across the organization the parent contracts folder can be targeted. If the interest is focused on only the sales contracts of a single division within the organization, over-collection can be avoided by targeting just the sub-folder for that division.

A sophisticated data collection approach should also make it easy to gain insight into what the data indicates. With pre-collection analytics, users can rapidly understand the scope of the data, and advanced search functions should allow targeting of specific datasets within an endpoint, network or cloud source, based on keywords, hash values or metadata. Further, collections should be able to be conducted in parallel including the ability to split jobs by folder so data can be analyzed sooner as individual jobs are completed instead of waiting for entire processes to finish.

opentext™

"Simply put, speed matters in corporate fraud investigations. The days of five-year investigations, of agreement after agreement tolling the statute of limitations—while ill-gotten gains are frittered away and investor confidence sinks—are increasingly a thing of the past."

Acting Principal Deputy Assistant Attorney General Trevor McFadden

Learn more

Discreet operation

Disruption shouldn't play any part in modern data collection. Data collection should run quietly in the background without monopolizing system resources or preventing the organization from continuing with its work by bogging down routine tasks

Failsafe measures

With dispersed data, sporadic connectivity is a fact of life. Data collection solutions should maintain logs of their successful processes and automatically reattempt any collection that fails, such as when a device drops off of a network. EnCase Information Assurance, and other forensic data collection solutions, communicate with remote agents to monitor connection attempts and automatically execute retries until devices re-appear on networks and collections are completed

Defensible process and results

Data collection is worthless 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 legally sanctioned output formats such as EnCase Information Assurance's LEF (Logical Evidence File) that maintains the integrity of collected data including that metadata has not been altered.

Easily transferable outputs

One of the core objectives of data collection is to enable subsequent data review, so data must be collected in an industry-standard format that can be readily ported to widely used review platforms, such as OpenText™ Axcelerate™. The ability to easily transfer data collections to review platforms is an important element of containing the cost of eDiscovery. Minimizing the effort required to load data into review platforms goes hand-in-hand with culling data at the point of collection to minimize the volume of data passed forward for review. Easy uploads and minimized review sets combine to maximize eDiscovery cost savings.

opentext™

Conclusion

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

That approach should allow organizations to efficiently collect all types of data from all sources and repositories in a single, easy-to-use process. It must be efficient, culling data as it is collected to minimize the volume of data passed forward for review while enabling rapid insights into its scope. It should be discrete to not monopolize computing resources and should include failsafes to counteract network disruptions and disconnections. Finally, it must be executed in a forensically sound, defensible process that passes judicial muster and generates a widely accepted review-ready format.

Fulfilling these seven critical requirements is a best practices prescription to accommodate the new mandates of modern data collection.

About EnCase Information Assurance

EnCase Information Assurance, a forensic data collection and preservation platform, exceeds all these requirements. It enables rapid data collection from an extensive array of sources and endpoints, automatically reduces data volumes and thereby review costs, and allows users to gain rapid insights into the scope of a matter. Users can combine keywords, hash values or metadata properties to search across all content systems without pre-indexing. It operates discreetly without disruption to users and corrects for network connection issues. EnCase Information Assurance is trusted by courts and is cited or mentioned in over 100 U.S. judicial opinions and publications. EnCase Information Assurance exports can be readily ingested by widely used review platforms in a variety of formats, including Concordance, EDRM XML and E01. It generates output load files that are Relativity-ready and supports streamlined uploads to OpenText Axcelerate.

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