VERITAS

Going Beyond Microsoft Office 365 Compliance Center with Veritas

How Veritas eDiscovery Platform Meets All Your Compliance Needs and Simplifies the eDiscovery Process

Veritas™ White Paper | July 2020

Contents

OVERVIEW . . . 3

WHAT IS eDISCOVERY AND SUPERVISION/COMPLIANCE? . . . 3

What tools does an organization need to meet compliance and governance challenges? . . . 3

Microsoft Compliance Center . . . 3

Google Suite . . . 4

eDISCOVERY AND COMPLIANCE SOLUTIONS . . . 5

Must-Have Solution Features . . . 5

Document Viewing Options . . . 5

Redaction . . . 5

Additional Content Sources . . . 5

Audio Indexing . . . 5

Optical Character Recognition . . . 6

Email Threading . . . 6

Machine Learning . . . 6

Data Visualization . . . 6

ADDRESSING eDISCOVERY CHALLENGES . . . 6

The Veritas eDiscovery Platform . . . 7

Unified Collections . . . 7

Processing and Analysis . . . 7

Review and Production . . . 7

Legal Hold . . . 7

SUMMARY . . . 8

OVERVIEW

Today, organizations work in a very agile environment. Users increasingly leverage software as a service (SaaS) productivity toolsets to complete their workloads. Products like Microsoft Office 365 and Google Suite offer users feature-rich environments with a wide array of utilities to make their daily lives easier. These tools provide email, social media, productivity applications, collaboration and convenient storage mechanisms.

Although these are great toolsets for users, they create greater compliance and data privacy risks for companies. How do compliance, legal and security teams cope with these tools while regulations become more stringent? Both Office 365 and Google Suite offer basic utilities to help organizations meet compliance and governance challenges; however, these tools fall short of best practices and can be cumbersome to use.

WHAT IS eDISCOVERY AND SUPERVISION/COMPLIANCE?

eDiscovery is the process of seeking and finding relevant information in electronic format, typically in response to legal matters and investigations. Supervision/compliance is a similar function, but is usually reserved for financial institutions that are required to monitor communications between financial broker dealers and their customers.

What tools does an organization need to meet compliance and governance challenges?

The most common response to the question about what tools an organization needs is compliance and eDiscovery applications that are specifically designed to search for relevant data needed for the presentation of data for legal case matters or data privacy compliance requests.

Microsoft Compliance Center

Microsoft's approach is to provide administrators with the Compliance Center. This toolset can help legal and compliance teams complete a great deal of their work, but there are some caveats organizations should be aware of prior to deciding on a process for completing the required searches.

What if the indexing tool doesn't index the entire content of documents or doesn't index the type of document that needs to be searched? If everything isn't indexed, a search can't retrieve all the responsive data needed, which is obviously a problem. So, let's look at the documents that are indexed by the email system in Office 365 (see Table 1 from the Microsoft website).

(Note: To see the list of file name extensions for SharePoint Online indexed documents, refer to the Microsoft website.)

Filter	File extension
Email message	.eml
Graphics Interchange Format	.gif
JPEG	.jpeg
Microsoft Excel	.xls, .xlt, .xlsx, .xlsm, .xlb, .xlc, .xlsb
Excel File	odbcexcel
Microsoft InfoPath	.infopathml
Microsoft Office Binder	.obt, obd
Microsoft PowerPoint	.pptx, .pptm, .ppt, .ppsx, .ppsm, .pps, .ppam, .potm, .pot, .potx
Microsoft Publisher	.pub
Microsoft Word	.doc, .docm, .dotx, .dotm, .dot, .docx
Microsoft XML Paper Specification	.xps
OneNote	.one
OpenDocument Presentation	.odp
OpenDocument Spreadsheet	.ods
OpenDocument Text	.odt
Outlook Item	.msg
Portable Document Format	.pdf
Rich Text	.rtf
Text	.txt
vCalendar	.vcs
vCard	.vcf
Visio	.vdw, .vsd, .vss, .vst, .vsx, .vtx, .vssx, .vssm, .vsdm, .vstx, .vstm, .vdx
Web archive	.mhtml
Web page	.html
XML document	.xml
ZIP archive	.zip

Table 1. File Name Extensions Indexed by Office 365 (for Exchange/Exchange Online)

You'll notice that the items indexed by Office 365 are primarily Microsoft Office files, which covers a good number of the file types users create and transmit daily. In the world of text-based file types, however, this amount represents a small number of the actual text-based files currently in use. Compared to several products on the market that index over 500 file types, Microsoft's product clearly falls short—and presents an opportunity for us to help make it better.

You'll also notice that the PDF files in Table 1 are listed as indexable and we don't see any graphic images. Graphic-based PDF files can't be indexed by the Exchange Online engine and graphic images can't be processed by the index (using optical character recognition [OCR]), which again opens the door to miss important data. Many image-based PDFs are "forms" that can contain personally identifiable information (PII) or personal credit information, which are types subject to almost every data privacy regulation on the planet. These types of forms are used for things such as credit applications, loan applications, passport requests and insurance forms. Organizations that can't locate these documents could be at risk of violating current data privacy rules.

Now that we understand the importance of indexing, what happens when items haven't been completely indexed or aren't capable of being indexed by the index engine? In the case of the Office 365 suite, these are appropriately referred to as "Partially Indexed Items." A content search run from the Security & Compliance Center automatically includes partially indexed items in the estimated search results. (Note that doing so just highlighted data that is partially indexed.)

Here are some of the reasons why items can't be indexed and are returned as partially indexed when you run a content search in Office 365:

- Indexing limits—Partially indexed items are Exchange mailbox items and documents on SharePoint and OneDrive for Business sites that for some reason weren't completely indexed for search. Most email messages and site documents are successfully indexed because they fall within the indexing limits for email messages; however, some items may exceed these indexing limits and will be only partially indexed.
- Unsupported attachments—Email messages may have an attached file of a file type that can't be indexed; in most cases, the file type is unrecognized or unsupported for indexing.
- Image files—These file types require OCR capabilities for indexing.
- Invalid file name extensions—Email messages that have an attached file without a valid handler, such as an image files, can't be indexed; this is the most common cause of partially indexed email items.
- Multiple files—Email messages with too many files attached may not be indexed.
- Large attachments—Files that are too large and are attached to an email message may not be indexed.
- Indexing errors—Even when the file type is supported for indexing, an indexing error can occur for a specific file.

According to Microsoft's documentation, most Office customers have **less than 1 percent of content by volume (total number of files) and less than 12 percent of content by size (size of the files) that is partially indexed**. The size of files is most important because larger files have a higher probability of containing content that can't be completely indexed. Therefore, if your organization has 1 PB (petabyte) of data to search through, the content by size data partially indexed could be as high as 120 TB (terabytes) or 10 TB of content by volume partially indexed. This seems like a large amount of missed data that could create risk.

Google Suite

Google Suite provides a feature-rich set of tools with collaboration, messaging and ease of use at its core. However, Google Suite provides a limited toolset for compliance and legal teams to respond to today's demands. When it comes to capturing email, journaling is the gold standard messaging and legal teams have relied on for receiving immutable versions of the email stream that contains data that hasn't been altered or reviewed by the user. Journaling is a copy of all correspondence that flows through the message transport. Google Suite only provides the Journaling feature in its Enterprise License, creating an important concern. Other features that are missing or aren't scalable in Google Suite are true financial compliance and eDiscovery tools that meet the compliance and eDiscovery needs of medium to large organizations.

eDISCOVERY AND COMPLIANCE SOLUTIONS

With so many compliance and eDiscovery software options available to organizations today, it can be difficult to find the solution that best suits your company's needs. As the current Microsoft and Google solutions illustrate, features vary widely. One tool is feature-rich but is lacking in a few areas, and the other needs some serious help to provide organizations with the functionality they require. So, where do we go from here?

Although compliance and eDiscovery review tools come in either desktop or cloud-based versions, all review applications share a common set of core features, with some variations. In the Microsoft Compliance Center, for example, there's a set of tools that allows you to get the job done—to a certain point. But now, let's dive a bit deeper.

Must-Have Solution Features

When looking for a solution to help your organization fully comply with legal, compliance and data privacy needs, there are many features, functions and benefits to consider. The best way to choose a good solution for your organization is to become as educated as possible about the technology and carefully compare the options. Fortunately, a good tool will usually pay for itself in the first incident by making document review more efficient and helping you locate "the smoking gun" evidence in more of your cases.

All enterprise tools include the "big three" core features—recursive parsing, search and indexing and tagging and organizing documents. The more features a tool has, however, the easier the search, review and presentation of the data will be. So let's dive into some additional components that can be extremely valuable.

Document Viewing Options

When you open a document within the tool, some may display it as a chunk of plain text or similar to what the document would look like if opened in its original application. When reviewing documents, being able to view a styled version without opening the original document is extremely useful. Remember that opening the original document is rarely a good idea because you could inadvertently change the file or metadata and risk infecting your computer with a virus or malware.

Redaction

Redaction is the process of protection used to describe removal of some document content by replacing it typically with black rectangles that indicate what was removed. For example, originally classified documents released under the Freedom of Information Act (FOIA) may include sensitive information redacted in this way. This practice is covered later in this white paper under an alternative name—. eDiscovery and compliance tools that offer this feature can also include options such as find and redact, which helps users to avoid missing information that must be redacted.

Additional Content Sources

The content referenced in both Office 365 and Google Suite is the typical target for compliance and eDiscovery. However, there are many data types that are not included in these productivity suites. Other content sources like social media—Twitter, Facebook, Snapchat, LinkedIn—are also typical targets for litigation and compliance along with other important sources such as phone text messaging. Some tools can ingest, index and search over 80 different content sources including resources like WebEx, Skype and Box. The Veritas eDiscovery Platform, for example, can index hundreds of document types—far more than any other tool on the market.

Audio Indexing

Electronic document productions often contain a variety of media files, but you can only search them for matching text. Because you can't extract text from audio files, they are invisible to the search engine. Fortunately, some compliance and eDiscovery applications, including eDiscovery Platform, now offer the ability to index audio files, permitting those files to be searched with text-based queries. If audio indexing is not available, some review tools will flag them and other files that lack text content so you can review them manually.

Optical Character Recognition

OCR is to image files what text indexing is to audio—they both locate text in non-textual media. Sometimes document sets include images taken from a smartphone or camera, and those images could be pictures of documents or even screenshots. A tool needs to include OCR for the information in these images to be visible to the search index.

Email Threading

Emails are some of the most common sources of electronic evidence—and can also be the most revealing. Having adequate email review functionality is critical for any document review project. Each email in the review tool should be grouped with its attachments along with other emails in the "chain" (thread) to which the email belongs. For example, if you are viewing an email that is a reply to an original message and that was subsequently replied to, there should be links to the original message as well as the later replies.

Machine Learning

Machine learning (also known as "predictive coding" or "technology-assisted review") is an exciting technology that offers the promise of locating relevant documents accurately and automatically by relying on patterns and inference to make decisions about what is relevant. Some compliance and eDiscovery experts believe machine learning can locate relevant documents faster and more accurately than keyword searching. When you need to work on a case with a large data set consisting of multiple terabytes of electronically stored information, choosing a platform with machine-learning capabilities is advisable. Better yet, choose a system that provides intelligent review that includes automated decision-making based on a reviewer's previous outcomes.

Data Visualization

Data visualization runs the gamut from simple pie graphs showing the relative frequency of various file types in the source data to complex diagrams showing the volume of email traffic between different persons of interest in a case over time. Although well-designed data visualization can help litigators see the entire forest and not just the individual trees, visualizations are often gimmicky and not helpful. It's important to have an idea of what you'd like to see depicted in visualizations first, and then assess whether a specific tool meets that need.

ADDRESSING eDISCOVERY CHALLENGES

No discussion of eDiscovery is complete without an understanding of the eDiscovery process. The Electronic Reference Discovery Model (EDRM) is the foundational workflow that encompasses all the necessary steps for successful search and response procedures (see Figure 1).

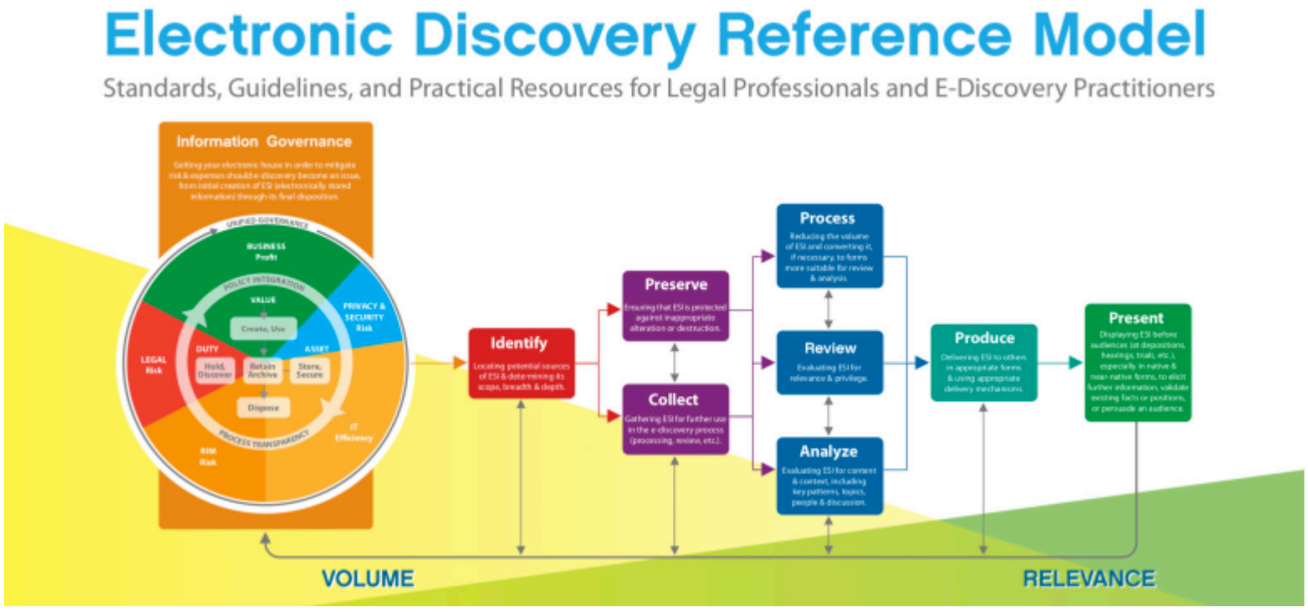

Figure 1. An overview of the EDRM workflow.

The Veritas eDiscovery Platform

Veritas has worked closely with our customers, their investigators and litigators to understand the serious challenges they face regarding compliance, investigation and litigation issues. Because many of our customers have faced similar issues, we have incorporated the most common ones into our eDiscovery solution—the Veritas eDiscovery Platform. The eDiscovery Platform is an end-to-end solution that encompasses the entire eDiscovery process and provides tools that meet all your compliance, FOIA and California Consumer Data Privacy (CCPA) needs in a single, easy-to-use, web-based application.

The four modules in the eDiscovery Platform support the entire eDiscovery lifecycle. Let's take a closer look at each of these components.

Unified Collections

The eDiscovery Platform's collections module lets you collect data from various content sources through a single user interface. You can choose either to have your IT department perform the collections or allow your legal compliance or investigation team do the collections, totally bypassing your IT department. This approach removes the collections burden from IT, freeing them to focus on their core responsibilities. It also gives the legal compliance or investigation team full control over where and from whom to collect data, where to send the collected data and how to manage the process. The collections module identifies custodians (the subject of the search), has a desktop/laptop search and a collection tool that builds an interactive map of custodians and their data sources, collects to a preservation store (legal hold) and filters data by keyword and metadata. Because the module collects data from multiple sources and reports on what was collected and when, it allows your organization to collect data in a defensible manner.

Processing and Analysis

The next phase in the EDRM model is the processing of the data. During this phase, the eDiscovery Platform's processing and analysis module enables rapid and accurate filtering, processing (indexing), searching and data analysis in multiple formats and languages. Earlier we looked at the number of items indexed in Exchange Online and Google Suite, which amounts to only a handful of document types. In comparison, the eDiscovery Platform can index hundreds of document types and also perform OCR on images, process audio and video files for search and categorize data through multiple policies, making search much easier. Using this module, corporations, government agencies and law firms can perform early case assessments and rapidly cull down data, reducing overall eDiscovery costs. The module also supports the iterative workflows required during real-world eDiscovery. Most important, it delivers deep insight into case facts and enables a new level of transparency and defensibility throughout the eDiscovery process.

Review and Production

The eDiscovery Platform's review and production module accelerates the review process. It provides unprecedented scalability and introduces the flexibility to deploy case-dependent linear review and predictive coding and intelligent review, expediting document review. This module also eliminates the need to create load files and enables an iterative electronic discovery workflow. Other features included in this module are concept-based search (find the secret), audio and video keyword search, document redaction (bulk redact, find and redact and other important redaction features), conversation threading (learn who sent what to whom), item tagging and machine learning for technology-assisted review. The full-featured toolset allows an organization to produce data in multiple formats for court presentation, hand off to opposing counsel or turn over for data privacy and FOIA requests.

Legal Hold

The eDiscovery Platform's legal hold module streamlines and automates legal hold management. It enables legal teams to satisfy the duty to preserve from anticipation to completion of litigation by providing a repeatable workflow. The module lets you manage hold notices and rapidly identify and collect critical data on demand. Among the features of the module are hold notices, reminders, escalations, custodian survey creation, a custodian portal (to answer requests) and a complete reporting structure. The legal hold module minimizes the risk of sanctions while providing the highest level of defensibility across the entire eDiscovery lifecycle.

SUMMARY

Unlike the Office 365 Compliance Center and Google Suite tools, the Veritas eDiscovery Platform is a standout in the industry. eDiscovery Platform provides a simple user interface with remarkable tools that make the eDiscovery process as easy as placing an order from Amazon. Whether you're looking for the smoking gun to prove your case or fulfilling a time-sensitive compliance request, eDiscovery Platform is the tool of choice for leading organizations, legal teams and compliance officers.

ABOUT VERITAS

Veritas Technologies is a global leader in data protection and availability. Over 50,000 enterprises—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.

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

For specific country offices and contact numbers, please visit our website.
www.veritas.com/company/contact

V1074 07/20