Office 365 meets evolving eDiscovery challenges in a cloud-first world
Office 365 meets evolving eDiscovery challenges in a cloud-first world

Contents

Executive Summary 1
Introduction 2
  Key benefits of Office 365 eDiscovery 2
  In this paper 3
Compliance and eDiscovery in Office 365 3
  Managing eDiscovery 3
    Prerequisites for eDiscovery 5
    eDiscovery security roles 5
Importing content into Office 365 6
Collecting and preserving content 7
  Advantages of eDiscovery in-place hold in Office 365 7
  How in-place hold works in Exchange Online 9
  Advantages of in-place hold over collection and email journaling 10
  How in-place hold works in SharePoint Online and OneDrive for Business 11
Using Office 365 eDiscovery search to filter content 11
  Exporting search results 12
Using Advanced eDiscovery 12
  Preparing source content for Advanced eDiscovery 13
  Using analytics to organize and reduce the volume of content 13
  Using predictive coding to determine relevance 13
  Validating the results of predictive coding 14
  Analyzing results quickly 14
  Remediating content using Advanced eDiscovery 14
  Exporting content from Advanced eDiscovery 14
Auditing eDiscovery events 15
How we use eDiscovery at Microsoft 15
  Placing content on hold 15
Using Office 365 eDiscovery search to filter content 17
Automating search and export 18
Validating search results 20
Using Advanced eDiscovery at Microsoft 20
Managing small-scale internal review projects 21
Our key takeaways for eDiscovery 21
For more information 21
Microsoft IT Showcase 21
Executive Summary
To continue to meet legal, business, and regulatory compliance challenges, businesses must be able to keep and protect important information and quickly find what’s relevant. Spending days, if not weeks, manually sifting through millions of files to find the small number that are relevant isn’t just expensive, it isn’t an option.

This paper walks you through the eDiscovery capabilities in Microsoft Office 365 and gives examples of how we use them at Microsoft to help satisfy compliance and legal requests in a timely and cost-effective manner.

When organizations migrate to the cloud, they are better served by solutions that are designed for the cloud from the beginning. That's why at Microsoft, we've adopted a cloud first strategy. Our solutions give our customers increased efficiencies, cost savings, and security in the cloud, right from the start. Our Office 365 eDiscovery solution brings eDiscovery to the cloud in a scalable, efficient, always up to date, and secure environment.

Office 365 eDiscovery can help you quickly and cost-effectively locate, identify, and retrieve relevant information—and preserve it in place. No need to move content to a separate archive to store, index, and process. And the Office 365 eDiscovery solution is available globally to use in any locale or situation where you need to respond to legal and compliance needs or to an internal investigation.

Complementing the eDiscovery capabilities is Office 365 information governance. It helps preserve the content you need and eliminate what you don’t, minimizing over-preservation and reducing the risk and expense of eDiscovery, investigations, and regulatory compliance.

When you need to respond to a legal or regulatory information request, the search and analytics tools in Office 365 eDiscovery can cut your costs and streamline your responses. eDiscovery search finds text and metadata in content across your Office 365 assets—SharePoint Online, OneDrive for Business, Skype for Business Online, and Exchange Online. Office 365 Advanced eDiscovery further organizes and filters your content. It groups content into categories, removes duplicates, and uses machine learning to filter for relevance, reducing the amount that must be sent to review. You'll find relevant content faster—while keeping your organization's information more secure.

At Microsoft, Office 365 eDiscovery is saving us about $4.5 million annually by reducing the manual work required for discovery requests.

At Microsoft, we know how demanding and complex compliance it can be. As you might imagine, being a large enterprise operating at a global scale, we’re subject to many discovery requests every year. Our legal department uses the eDiscovery features of Office 365 to improve the accuracy and usefulness of our discovery results and save time and money.

Before Office 365 eDiscovery was available, we had to manually collect content from various sources. Gathering a large volume of content and loading it into an offline processing tool took time. Then we had to reprocess it. With collection, processing, and remediation, it could take between two and three weeks to give outside counsel the documents they requested. Today, we do most of this work in hours, not days or weeks. We start to export content on the fly and have it ready for counsel to load into their review tool by the end of the day.

When we need to find specific content to respond to discovery requests, we first use eDiscovery search in the Office 365 Security & Compliance Center. We run searches right away, across the relevant Office 365 assets, without requiring the preliminary step of collecting content and moving it to a separate location to index and search.

We also preserve relevant content in place, in Office 365. We associate the relevant content sources with a case that we create in the Security & Compliance Center and then place the content on hold. This hold overrides any other retention policies that might be in force, and preserves the content for the duration of the case. The hold is practically invisible to the people using the sources, so they can continue working on their projects without interruption or loss of productivity.

After we discover potentially relevant content using Office 365 eDiscovery Search, we use Advanced eDiscovery analytics to thread email conversations, remove duplicates, find near-duplicates, and identify themes. This lets us give each reviewer a structured batch of unique files, eliminating redundant effort and saving review time. In some cases,
Instead of doing heavy keyword culling, we use the Advanced eDiscovery Relevance feature to identify relevant content. And even if we’re using keyword filtering, we always use Advanced eDiscovery to export our content in a format that’s immediately usable by our eDiscovery review partner and which requires no reprocessing.

By reducing the amount of manual work required to respond to eDiscovery requests, Office 365 eDiscovery saves our legal department about $4.5 million annually. With eDiscovery search, we typically reduce the amount of content in a case by about 95 percent. However, this still leaves large volumes of data that need to be submitted to the very costly process of legal review. Advanced eDiscovery helps us reduce these costs significantly: we typically see a further reduction of 30 percent by eliminating duplicate files and grouping near-duplicates, and another 25 percent by consolidating email threads.

**Introduction**

It can happen in any organization. You’re going about your business, and you receive a discovery request for “any and all” information (email, documents, presentations, databases, instant messages, images, voice mail, social media posts, and so on) related to a project you completed last year. Or you need to collect content that demonstrates you’re complying with corporate or government rules. If only the tools required to manage those tasks were just built into the platform where the data is. Fortunately, when your data is in Office 365, they are.

To be prepared for internal investigations, external litigation, or regulatory requests, your organization needs to preserve potentially relevant content. At the same time, you want to find relevant content quickly without disrupting your business. Preserving content you don’t need impedes your ability to do this and increases your overall risk.

We have seen that as businesses grow, so too do the demands to be compliant. The Office 365 Security & Compliance Center provides the solution. Its compliance features help you protect important content and reduce the expense and risk of keeping content you don’t need. And its eDiscovery features make it easier to identify content that’s relevant to a specific investigation, preserve it, and get it ready for a requesting party or reviewer.

Whether you are a small business or a large enterprise, the complexities of compliance are simplified with Office 365. Small businesses can grow quickly and achieve compliance with a single step. Large enterprises will find complex compliance requirements simplified and advanced capabilities just a click away.

**Key benefits of Office 365 eDiscovery**

As organizations migrate to the cloud, they need solutions designed for the cloud from the start, not simply older tools that have been shoe-horned into this new environment. That’s why our cloud first strategy requires that we build new solutions that give our customers increased efficiencies, cost savings, and security in the cloud. eDiscovery has traditionally been on premises where information is manually collected from various sources and processed to find the most relevant data. Our Office 365 eDiscovery solution brings eDiscovery to the cloud in a scalable, efficient, up to date, and secure environment.

Office 365 eDiscovery offers many benefits, including:

- **Global availability.** The Office 365 eDiscovery solution is available globally to use in any locale or situation where you need to find and access content to respond to legal and compliance needs or to an internal investigation.

- **Cost savings.** Office 365 eDiscovery helps you identify the most relevant content more quickly and easily, with far less manual review than previously possible. In legal matters, with less content to send to third-party reviewers, review costs are significantly reduced.

- **Faster responses to eDiscovery requests.** Content that you place on hold in Office 365 is preserved in place. You don’t need to move it to another archive for preservation and then wait for it to be indexed before you can search it. Office 365 eDiscovery lets you quickly identify and export relevant content when you need it.

- **Less manual work.** Enhanced remediation in Advanced eDiscovery reduces the need to manually remediate unsearchable content. Also, the ability to port relevant content directly into third-party review tools eliminates the need for manual processing to enable ingestion.
In this paper

The first section of this paper, Compliance and eDiscovery in Office 365, introduces information governance in Office 365 and the Office 365 eDiscovery features and workflow. The second section, How we use eDiscovery at Microsoft, describes how the Microsoft legal organization uses Office 365 eDiscovery features and lists some key takeaways that the team has learned.

Compliance and eDiscovery in Office 365

Today’s organizations face information overload. The amount of electronic information is exploding. And the information is more complex, coming from multiple sources in multiple formats—email, documents, social media, instant messages, and videos—the list goes on.

Managing information effectively to meet internal and external compliance requirements is more difficult than ever. The solution starts with effective information governance. In Office 365, we deliver cloud-powered, intelligent, in-place information governance solutions that address importing, retaining, protecting, and purging files when a scheduled expiration date occurs. These solutions help you keep important information; delete what’s redundant, obsolete, or trivial; and manage how sensitive or confidential information is shared. The high-value, important content in your organization can be protected for as long as you need it.

Office 365 also gives you flexibility in the way that you preserve important content. You have the option to preserve content at the global level, using organization-wide policies, or at the eDiscovery case level, in relation to a specific investigation. You can apply a preservation policy globally to certain content to preserve it regardless of events, or place content associated with a case on hold, preserving it for the duration of an investigation.

Managing eDiscovery

The Electronic Discovery Reference Model (EDRM), shown in the following figure, summarizes the typical phases in the eDiscovery process for identifying relevant content and reducing the volume of content to present.

![Image](https://edrm.net/)  

*Figure 1. Electronic Discovery Reference model. Source: edrm.net*  

---

1 Image source courtesy of edrm.net. The source image was made available under Creative Commons Attribution 3.0 Unported License.
The information governance features in Office 365 help you intelligently manage content in a proactive manner to respond to both internal and external compliance requirements. You can respond to eDiscovery requests more quickly, easily, and cost-effectively. The Office 365 Security & Compliance Center is a central location where you manage information governance and eDiscovery across all of your Office 365 data assets. Use it to:

- **Import content into Office 365 so that you can manage it.** For on-premises content stored in Exchange or file shares, Office 365 provides an import service. For external content from social and messaging apps, document collaboration tools, and vertical apps (such as CRM or financial sites), some Microsoft partners provide connectors that support different third-party file formats.

- **Apply information governance to Office 365 content.** Information governance helps ensure that content is managed in a way that supports your compliance needs. It lets you proactively set retention policies on your Office 365 content in mailboxes, public folders, and sites. You can preserve important content and delete content when you no longer need it. Or for specific investigations, you can preserve just the content associated with an eDiscovery case by putting the content on hold. This hold overrides any retention policies that would otherwise apply.

- **Create and manage eDiscovery cases.** eDiscovery cases facilitate investigations and restrict access to them. When you create a case, you can manage it as follows: add members and give them specific permissions to control the types of actions they can perform in an investigation; place content source locations associated with the case on hold; preserve on-hold content indefinitely, until you remove the hold, without saving content to a separate archive; specify a date range and keywords to narrow the content that’s preserved; run broad or targeted queries; prepare content for deeper processing and analytics with Advanced eDiscovery; and export content for review and production.

- **Preserve content in place.** On-hold content remains in place, where it’s located in Office 365, so that custodians—the content creators—can continue to access it. The content—email, messages, calendars, and files stored in Exchange Online, SharePoint Online, and OneDrive for Business—remains accessible to custodians, without duplication or stubbing. When you preserve content in Office 365, it’s preserved in an immutable form. If someone modifies or deletes on-hold content, the system preserves the original version in a secure location. Office 365 preserves only the content that the policy or custodian has tried to delete, so there’s no duplication. It saves one definitive copy, and you don’t use additional disk space for extra copies.

- **Run simple or complex search queries.** Office 365 eDiscovery search lets you query all of your organization’s Office 365 content, or focus searches on particular content source locations used by relevant custodians. To search efficiently, you don’t need to copy your content to a consolidated external repository. Search all mailboxes and public folders in place in Exchange Online, all SharePoint Online sites, and all OneDrive for Business source locations in a single search.

- **Quickly analyze content for relevance.** Advanced eDiscovery helps you analyze large data sets and find content that’s most relevant to a case. It can also organize your content, making the legal review process easier and more efficient.

- **Export content for review.** You can export files from Office 365 search results in their native format. Advanced eDiscovery additionally exports content in a format that third-party review applications can directly ingest.

- **Audit activities in Office 365.** You can monitor security and bolster the defensibility of your eDiscovery results by logging user and administrator activity in Exchange Online, SharePoint Online, and OneDrive for Business. Log files are stored for 90 days in Office 365. For longer term storage, download the log files using the Management Activity API. View reports in Office 365 Security & Compliance Center, or create custom reports by using the Management Activity API.

**NOTE:** The Security & Compliance Center is fully scriptable using PowerShell, enabling you to manage your Office 365 Security & Compliance Center settings from the command line. For more information, see [Office 365 Security & Compliance Center PowerShell](microsoft.com/ITShowcase).
Office 365 meets evolving eDiscovery challenges in a cloud-first world

The following figure illustrates the information governance and eDiscovery features of Office 365.

Figure 2. Information governance and eDiscovery in Office 365

Prerequisites for eDiscovery
To use the eDiscovery features of the Security & Compliance Center, you need the following:

- **Office 365 eDiscovery.** If you want to perform eDiscovery searches to filter content sources for potentially relevant content, your organization needs to have an Office 365 Enterprise subscription. An E3 subscription gives you the ability to search, and an E3 subscription lets you create case holds and export search results.

- **Office 365 Advanced eDiscovery.** If you want to analyze a custodian’s content using Advanced eDiscovery, the custodian must be assigned an Office 365 E5 license. Alternatively, you can assign an Advanced eDiscovery standalone license to custodians who have an Office 365 E3 license. Administrators and compliance officers who are assigned to cases and use Advanced eDiscovery to analyze content don’t need an E5 license.

eDiscovery security roles
When you create a case in the Security & Compliance Center, you assign roles to the people whom you’re going to add as members to a case. This enables them to do specific eDiscovery tasks.

- **eDiscovery Manager.** A person assigned to this role group can create and manage cases, run searches, preview results, and export results. This person can’t access or manage cases that other eDiscovery managers have created or see what they’re searching for, unless he or she is specifically added as a member to those cases. By default, this user can search across all Office 365 content for the organization without needing specific access to each site (although you can limit access by using compliance security filters, as described a little later). This user has read-only access to the content through the eDiscovery tools, but they can’t access the content as an ordinary user would, such as through a browser. eDiscovery Managers also perform administrative tasks in Advanced eDiscovery, such as creating cases and importing data.
Office 365 meets evolving eDiscovery challenges in a cloud-first world

- **eDiscovery Administrator.** A person assigned to this role group is also a member of the eDiscovery Manager role group and has the same permissions. An administrator can access all eDiscovery cases in the Security & Compliance Center, without being a member. eDiscovery Administrators perform administrative tasks in Advanced eDiscovery, such as setting up users.

  NOTE: Exchange and SharePoint administrators do not have these permissions unless they’ve explicitly been given this role.

- **Reviewer.** A person in this role group has a read-only view of content for cases this person is a member of. This role is generally used for Advanced eDiscovery investigations and assigned to outside counsel. It provides a good way to collaborate and let counsel view the results of an investigation while it’s in progress. It also lets outside counsel view the content in order to train the Relevance system in Advanced eDiscovery for that case. (This role can view only the content in the case to which they’re assigned.)

For details about these roles and instructions for assigning them, see **Assign eDiscovery permissions in the Office 365 Security & Compliance Center**.

In addition to assigning roles, you can use PowerShell cmdlets to set permissions filters that allow individuals to perform content searches on a subset of content sources based on your organization’s structure. If necessary, use these filters to determine who can search specific peoples’ content, instead of giving all eDiscovery Managers and eDiscovery Administrators the ability to search everyone’s content. For more information, see **Configure permissions filtering for Content Search**.

  NOTE: When you create a case, you’re automatically added to the case as a member. You can add other members who will also work on this case, as described in **Create a new case and add members**. Only case members can view the search results associated with a case.

**Importing content into Office 365**

To use the Office 365 retention and eDiscovery features with your content, the content must be stored in Office 365. For content outside of Office 365, we make it easy for you to import it. Upload email, documents, and other content to Office 365 to a network storage location in the Microsoft cloud, or put your content on an encrypted hard drive and ship it to Microsoft. Then use the Office 365 Import Service to import it. For more information and links to step-by-step instructions, see **Overview of importing PST files and SharePoint data to Office 365**.

  NOTE: Skype for Business content saved in Exchange Online doesn’t need any action to be searched by Office 365 eDiscovery. If your users turn on OneDrive for Business sync, their synced desktop content will be accessible to Office 365 eDiscovery features for indexing, search, analysis, and in-place preservation.

All types of content—not just email and documents—are potentially important and might be used in an investigation. Your content may come from social media, messaging, vertical industries (such as CRM or financial services), or collaboration tools like Dropbox. To import this content into Office 365, use a solution from a **Microsoft partner**. If you have specialized on-premises deployments of Skype for Business or Lync, our partners can also help import content from those deployments as well as Yammer content. Partners understand the file formats and can extract...
collecting and preserving content

An important eDiscovery task is collecting and preserving potentially relevant content, so that it can’t be changed or deleted. Historically, collection has been a prerequisite for preservation. You would locate the content and then move it to a separate archive. With Office 365, however, you don’t need to collect potentially relevant content before you can preserve and process it.

You have two ways to preserve content in the Office 365 Security & Compliance Center: either using retention policies or case-specific eDiscovery holds. Use retention policies to proactively manage the lifecycle of content globally in Office 365. Use case-specific holds to retain only the content associated with a specific investigation.

To use a case-specific hold, create a case in the Security & Compliance Center, create a Hold within that case, and add the relevant content source locations to that Hold. You can add sources such as Exchange Online mailboxes, OneDrive for Business sites, SharePoint Online sites, and Office 365 groups to the Hold. If you want, you can also enter a search query to target the Hold to specific content within those sources. After the Hold policy is applied, the data in the source locations will be preserved in place—both content that already exists as well as content yet to be created.

To identify where the content that you want to preserve is located, you may want to consult the custodians whose content is relevant to the matter at hand. They may provide a list of pertinent locations, such as their Exchange mailboxes, OneDrive for Business and SharePoint sites, local computers, file shares, other databases, line-of-business applications, social media, or others. You can also run a search across your organization’s assets, including SharePoint Online sites and Exchange Online, and then use the list of source locations to find the top locations—those yielding the most search hits.

Advantages of eDiscovery in-place hold in Office 365

The advantages of in-place hold in Office 365 are:

- **Saves time.** Content doesn’t need to be transferred out of Office 365, so you don’t have to spend time collecting, exporting, and transferring content to a third party and waiting for it to be indexed separately.
Reduces risk. Content isn’t duplicated or transferred to another provider or compliance boundary.

Yields higher content fidelity and lower costs. Content stays in Exchange Online and SharePoint Online, which results in lower storage costs and higher fidelity of content and metadata.

Lets you preserve content according to location, query, or policy. Preserve a mailbox or SharePoint site, apply a query to hold less content, or use preservation policies. Place a hold that covers all content within a source location, including content that hasn’t yet been created. Or, refine the hold by using queries with complex keywords and other metadata filters, like date, content type, domain, and participants.

Doesn’t impact custodians. Even though Office 365 content that you place on hold on is preserved in place, custodians can still work on it without disruption. This improves custodians’ productivity and their ability to collaborate. Custodians create, edit, and delete content without worrying about how it’s being preserved.

Preserves even deleted content. When a custodian deletes content, or if it’s due to be deleted under a policy, the system moves the content to a location that Office 365 eDiscovery searches can still find. To the custodian, such content appears to be deleted—yet behind the scenes, it remains intact, still indexed and discoverable by eDiscovery searches. When the hold is lifted, the original policy or user deletion takes effect.

Saves storage. Office 365 preserves only content that a policy or custodian has tried to delete, so there’s no double-saving. This gives you one definitive copy, and saves storage.

Content that isn’t needed isn’t preserved. Rather than suspending automatic deletion mechanisms to preserve content, content that doesn’t meet the eDiscovery criteria need not be placed on hold. It can continue to be subject to your organization’s normal retention policies. This helps keep custodians’ mailboxes from becoming overfull.

For step-by-step instructions about putting content on hold, see Place mailboxes and sites on hold.
How in-place hold works in Exchange Online

The following figure shows the structure of an Exchange Online mailbox. The top section, containing Inbox, user-created folders, and Deleted Items, is visible to the custodian. The Recoverable Items partition, however, isn’t exposed to the custodian, with one exception described later. When you place an Exchange Online mailbox on hold, Office 365 eDiscovery uses the Recoverable Items partition to preserve all of the mailbox content (email and attachments, tasks, and contacts) intact, regardless of what the custodian does with them.

Figure 4: Structure of an Exchange Online mailbox

Here’s what happens to Exchange Online content that’s been put on hold, under different scenarios:

- **A message or attachment is edited.** If a custodian edits a message or attachment, the edited version stays in the portion of the mailbox the custodian can see, and the original message is moved into Versions, completely intact and in an immutable state.

- **A message is purged.** If a custodian whose content is on hold deletes an item in Deleted Items, the item moves to Deletions in the Recoverable Items partition. If a retention policy deletes an item, the item also goes to Deletions. If the mailbox is on hold, when the number of days specified by the Single Item Recovery setting has expired (14 days by default), the item moves to Purges for the duration of the hold. When the hold is lifted, the original retention policy takes effect and the item is expunged—completely deleted from the system.

- **A custodian recovers a purged message.** If an IT department has configured this, people who use Exchange Online can recover purged items within 14 days by default. So, while they can’t normally see items in the Deletions folder, when recovering a purged item, they can. When a custodian’s email content is on hold, if a custodian tries to delete an item in Deletions, the item moves into Purges, where it continues to be preserved according to the Single Item Recovery limit and hold settings.

- **A message that’s on targeted hold is purged.** When a targeted hold has been placed on specific messages or attachments, a cleanup process runs roughly every week, which acts on items that exceed the Single Item Recovery limit. The cleanup process moves items that match the targeted query criteria to DiscoveryHold for
ongoing preservation. Items with any index errors are also moved to DiscoveryHold for continued preservation. The cleanup process then expunges the rest of the items that don’t match the query criteria and have no index errors.

- **A custodian archives Skype for Business content.** When a custodian sets their local client to archive Skype for Business content, the content is saved in the Outlook Conversation History folder. Content includes instant messages, received files, and meeting content. When an eDiscovery manager searches a custodian’s Exchange mailbox, Conversation History content is included in the search. When a custodian’s Exchange Online content is put on hold, their Conversation History content is automatically preserved, including any Skype for Business content it contains.

- **Exchange Online archiving has been set up for a custodian.** If a custodian or administrator has configured Exchange Online archiving, when message data reaches a certain age, such as 90 days, the system moves it into an archive repository in the cloud that the custodian can access by using Outlook or Outlook Web Access. When a custodian’s Exchange mailbox is placed on hold, the hold takes effect for this archive the same as for the main mailbox. If the custodian tries to delete archived content that’s on hold, the content is preserved in the Recoverable Items partition of their Exchange Online mailbox. Office 365 eDiscovery searches include the Exchange Online archive, so you don’t have to include it as a secondary source location. Every time you search a custodian’s mailbox, the archive is also searched.

**Advantages of in-place hold over collection and email journaling**

Historically, people have performed eDiscovery by using a collection system that gathers content in response to specific litigation or compliance events, or a journaling process that saves all content to a separate offline archive. When an email is transmitted, a copy of it and any attachments are delivered to the archive.

These approaches have several disadvantages:

- The collection approaches are time- and resource-intensive. All the potentially responsive content must be gathered, downloaded, loaded into an offline environment, indexed, and searched.
- Journaling creates a duplicate of the content, increasing costs due to the need to store a copy of the content.
- The journaling process misses items that were never sent or received, such as drafts and calendar appointments. It also fails to capture the full fidelity of certain metadata, such as the message read/unread status or the folder into which the custodian filed each item.
- In all offline approaches, content may be compromised while moving from one environment to another.
- Capturing content at a point in time and transferring it to an offline repository misses later edits and additions. This means that the content is never fully up-to-date.

An Office 365 in-place hold has distinct benefits:

- Email content isn’t moved to a separate repository for preservation. It remains in the custodian’s Exchange mailbox and is managed as described in the previous section.
- Office 365 indexes everything in Exchange mailboxes, including content in Recoverable Items. This makes preservation simpler to manage than email journaling, and it retains the structure of the content and metadata much better.
- A hold in Office 365 eDiscovery preserves all content in the mailbox—tasks, notes, contents, and so forth—and not just email messages.
- When a custodian searches their mailbox, only items in the portion of the mailbox that the custodian can view are returned. When an eDiscovery manager searches an Exchange mailbox, the search returns everything.

**NOTE:** The quota for the Recoverable Items folder is automatically increased to 100 GB when you place a mailbox on hold. If the Recoverable Items folder reaches or exceeds the 100 GB quota, you can increase the available quota for recoverable items by enabling the custodian’s archive mailbox. For more information, see [Increase the Recoverable Items quota for mailboxes on hold](microsoft.com/ITShowcase).
How in-place hold works in SharePoint Online and OneDrive for Business
SharePoint Online and OneDrive for Business similarly have a partition that isn’t visible to the user, the PreservationHoldLibrary. When a custodian’s content is on hold, deleted content is stored in this partition.

Here’s what happens under different scenarios with SharePoint Online and OneDrive for Business content that has been placed on hold:

- **On-hold content is due to expire under a retention policy.** You can apply retention policies to SharePoint Online and OneDrive for Business content. If content expires under a retention policy while a hold is in force, the content is preserved in PreservationHoldLibrary. When the hold is lifted, the content is expunged—completely deleted from the system.

- **Content is deleted.** Deleted content is preserved in PreservationHoldLibrary. When the hold is lifted, the content is expunged.

- **Document versioning is enabled.** If document versioning is enabled in SharePoint Online, when a custodian creates a new version of a document, the previous version is preserved in the site’s Versions directory.

Using Office 365 eDiscovery search to filter content
You can run eDiscovery searches to find and filter content that’s associated with a case. Office 365 indexes most Exchange Online, SharePoint Online, and OneDrive for Business content. This includes Office files, searchable PDF files, lists, communications, social discussions, and many other file types. You can use keywords and metadata filters to minimize how much content must be reviewed in your eDiscovery case, or use search to bring back entire sources without filtration.

Office 365 eDiscovery search supports typical legal queries. You can construct queries using Boolean and proximity operators to filter content with any combination of keywords, date ranges, authors, recipients, domains, file types, etc. You can apply a query to all of the content in a case, or narrow the scope of the query to a subset of source locations.

Among other attributes, Office 365 indexes Exchange message properties including sender, recipients, message body, and attachments. Documents and messages encrypted with RMS and Azure RMS technology are also indexed and searchable. For details, see [File formats indexed by Exchange Search](file formats) and [Default crawled file name extensions and parsed file types in SharePoint Server 2013](file formats).

Some subsets of content may not be indexed and can’t be searched, such as image files, password-protected items, or items encrypted with non-Microsoft technology. You can, however, use eDiscovery search to identify the items that have index errors, so you can export and remediate them as appropriate. And although such content may not be fully text-searchable in place, you can use metadata filters, such as date or email sender information. For details, see [Unindexed items in Content Search](unindexed items).

Very large eDiscovery searches can search all mailboxes, all Exchange public folders, all SharePoint Online sites, and OneDrive for Business source locations with a single query. If you don’t want to search everything, you can specify up to 1,000 mailboxes and 100 sites per query. There are, however, limits, including the number of results that you can preview, the maximum number of keywords in a single search (500), and the number of variants for wildcard terms (10,000 total). For details, see [Limits for Content Search in the Office 365 Security & Compliance Center](limits).

You can also specifically search items that were imported into mailboxes in Office 365 from a third-party source. For more information, see [Use Content Search to search third-party data that was imported into Office 365](use content search).

**NOTE: Search results associated with a case can be viewed only by case members who have been assigned the eDiscovery Manager role.**

After you run a search, the number of content source locations and an estimated number of search results are displayed in the details pane of the search page. You can preview the most recent 200 results per source location, up to 1,000 items per query, to help determine whether the search is appropriate or whether it requires refinement. To preview a document, click Preview Results and scroll through the presented results. After completing your search, you can export either a report or the full results to a local computer, or prepare the results for analysis in Advanced eDiscovery, as described later.
For more information about running searches, see Run a Content Search in the Office 365 Security & Compliance Center. For details about constructing queries, see Keyword queries and search conditions for Content Search.

Exporting search results
After running your search, you can export the results. You can opt to export either the full search results themselves, or simply a report that lists each item and its metadata, as described in Export a Content Search report. Or, if your organization has an Office 365 E5 subscription, you can apply Advanced eDiscovery analytics that further refine and organize the content before you export it, as described in the next section.

You can export the results of a single search or multiple searches. Results are exported as follows:

- Exchange Online content and Skype for Business content that's been archived in Exchange Online is exported in PST or MSG files. Encrypted Exchange items are exported with the encryption intact, however, in early 2017, we will add a feature to automatically decrypt RMS encryption upon export, further streamlining the process.
- SharePoint Online and OneDrive for Business content is exported in its native format.
- SharePoint Online pages are exported as MHT files.
- SharePoint Online lists are exported as CSV files.
- A manifest file (in XML format) that contains metadata information about every search result is included with the export.

After the native files exported from Office 365 eDiscovery are downloaded, they can be viewed using their native applications (such as Outlook or Word), but they may also be processed with a third-party tool in order to be loaded into a dedicated review tool. The content exported from Advanced eDiscovery, as described later, can be imported directly without any additional processing for most files.

For step-by-step instructions to export search results, see Export Content Search results from the Office 365 Security & Compliance Center.

Using Advanced eDiscovery
Advanced eDiscovery goes beyond search, using machine learning and predictive coding to intelligently analyze the content, organize it, and reduce it before it goes to review. It intelligently simplifies sorting through large quantities of content to quickly find what's relevant. It saves review time and costs—and gets you better results faster.

Advanced eDiscovery allows you to:

- **Organize content conceptually.** Advanced eDiscovery organizes content into themes, so that attorneys and investigators can quickly browse and discover key content.
- **Reduce the volume of information for review.** Advanced eDiscovery includes the Relevance application. An attorney who is familiar with the litigation at hand can train Relevance to automatically identify content that's relevant to the case.
- **Structure content for more efficient review.** Advanced eDiscovery organizes content by reconstructing email threads, identifying exact and near-duplicates, and intelligently batching content for review. These structures optimize the review process, replacing a traditional linear review by allowing attorneys to review documents in closely related and tightly structured groups.

The workflow for Advanced eDiscovery is as follows:

- As already described, you start by creating an eDiscovery case in the Security & Compliance Center and adding members to it.
- Then you identify potentially relevant content sources and optionally filter the content using Content Search.
- Next you analyze the search results with Advanced eDiscovery to further examine, sort, and reduce the volume of content.
• Finally, you export the analyzed results for review, either to a storage location or directly into a third-party review tool.

The intelligent analytics in Office 365 Advanced eDiscovery typically yield the following results:

- Grouping near-duplicate content reduced document review time and cost by between 20% and 30%.
- Grouping email into threads typically reduced the time and cost of email review by between 40% and 60%.
- Using Relevance resulted in a savings of between 60% and 90%.

Figure 5: Typical review time and cost savings using Advanced eDiscovery

Preparing source content for Advanced eDiscovery
You start with the source content from Office 365 eDiscovery search(es) that you’ve previously created for a case, which may or may not be filtered or refined. Then, follow the steps in Prepare search results for Office 365 Advanced eDiscovery. After you complete these steps, you can begin working with the content in Advanced eDiscovery.

Using analytics to organize and reduce the volume of content
Advanced eDiscovery includes analytics for structuring your content. This helps organize the content and reduce its volume. These analytics are useful and relevant even for smaller content sets, including those that have already been filtered using keywords. As noted briefly above, these analytical tools include the following capabilities:

- **Grouping near-duplicate content.** Advanced eDiscovery organizes items of content into groups of unique files, duplicates, and near-duplicates. This lets you structure your review more efficiently, so one person reviews a group of similar documents, and enhances consistency by having the same reviewer work on different versions of the same content.

- **Clustering content by theme.** Advanced eDiscovery analyzes content to identify and map salient themes. It creates a meaningful label to characterize each theme. Then it charts relationships between themes and documents. Viewing content that shares a theme can give you valuable insights beyond keyword search statistics.

- **Grouping email into threads.** In an email thread, later messages typically contain all the previous messages in the conversation. Advanced eDiscovery organizes email into hierarchically structured groups of email threads. This helps you identify the messages containing unique content and lets you focus on only the new information in each message. Email threading removes the need to review every message in its entirety in an email thread, eliminating review redundancy, and focusing your attention on the unique data in the thread.

For details about these capabilities, see Analyze case data with Advanced eDiscovery.

Using predictive coding to determine relevance
Predictive coding is a sophisticated way to refine content. Going beyond keyword search, the Advanced eDiscovery Relevance application uses machine learning to decide what’s relevant to the case and what isn’t. Relevance
consistently finds more relevant content, while returning less irrelevant content than conventional methods, such as keyword search.

When you use predictive coding, you (or someone familiar with the case or investigation, such as an attorney or subject matter expert) train the Advanced eDiscovery system to find the content you want. The system selects samples of documents which the attorney tags as relevant or not relevant to the case. The first tagging cycle builds a statistical model. The system uses the statistical model to monitor training progress and quantify results. Training the system includes a number of rounds (typically between 25 and 40, with each round comprising 40 documents) until the system has enough data to create a valid model for what you consider to be relevant or not. The system uses an Active Learning capability to select the training documents, maximize efficiency, and optimize outcomes. After training is complete, Relevance calculates the likely relevance of each document in the collection. It ranks the documents to help you decide which documents to submit to review based on relevance estimates.

For details about predictive coding, see Use the Advanced eDiscovery Relevance module.

Validating the results of predictive coding
Predictive coding has been used in many cases and has achieved broad acceptance in the legal community and by various government agencies as a valuable eDiscovery tool.

To help you demonstrate that you used predictive coding appropriately, Office 365 eDiscovery provides auditing capabilities and a decision log that shows all of the steps that you took to cull content. Part of the recommended workflow is to use the built-in statistical analysis to sample the set of irrelevant data. Sampling ensures that the data has been effectively and defensively culled. Many of our Microsoft eDiscovery partners can assist with the eDiscovery process and provide workflow advice and validation.

Analyzing results quickly
Two new features of Advanced eDiscovery make it faster and easier to find, analyze, and review relevant information:

- **Express Analysis.** With one click of a button, Express Analysis organizes your content into near-duplicates, email threads, and themes and then exports the results. No additional configuration is required.

- **Export with analytics.** Advanced eDiscovery exports a spreadsheet, Export_List.xlsx, that includes all of the content metadata and Advanced eDiscovery analytics information from your results. Hyperlinks open any item from the spreadsheet with a single click. And Advanced eDiscovery indicates when an item needs to be reviewed, or whether it contains only redundant information and can be defensively skipped. Getting your results in this spreadsheet format makes it easy to sort, filter, and annotate them, speeding up the review process.

For more information about these new features, see Accelerate your eDiscovery analysis workflow with one click.

Remediating content using Advanced eDiscovery
Most content in Office 365 is indexed and searchable, but there will always be some content that doesn’t have associated text. In many eDiscovery cases, most unindexed content is in image files. With its new optical character recognition (OCR) capability, Advanced eDiscovery will extract text from image files or objects within the files. The text can then be analyzed by Advanced eDiscovery analytics. This reduces the amount of manual remediation work required to analyze image files.

Exporting content from Advanced eDiscovery
After you complete your analysis and reduce the content set, you’re ready to export it from Advanced eDiscovery. When you export, all of the content associated with the case can be downloaded to your local computer or copied to another location. The content is exported in its native format with accompanying HTML representations and extracted text files, so the output is directly compatible with third-party review tools. The export package includes a load file in CSV format. The CSV includes the metadata of the exported content and all of the analytics metadata needed to organize the content, such as near-duplicates, threads, and themes. The export also includes the Export_list.xlsx file.
that you can use to organize and annotate your review of the downloaded content. As noted, the data, both native files and metadata, is structured for direct streamlined export to the leading review tools.

For more information, see Export case data with Advanced eDiscovery.

Auditing eDiscovery events
The Security & Compliance Center auditing features help ensure that your approach to eDiscovery is defensible in court. Use auditing to verify that on-hold content or its metadata wasn’t altered. Office 365 auditing captures a number of events for security and compliance monitoring. Content Search and eDiscovery-related activities that are performed in the Security & Compliance Center or run with the corresponding PowerShell cmdlets are logged in the Office 365 audit log. Events are logged when administrators or anyone who’s assigned eDiscovery permissions perform certain Content Search and eDiscovery-related tasks in the Security & Compliance Center. For details, see Search for eDiscovery activities in the Office 365 audit log.

How we use eDiscovery at Microsoft
At Microsoft, our litigation team uses Office 365 eDiscovery and Advanced eDiscovery to help save time and money in our investigations. We use eDiscovery to preserve and discover our potentially relevant content. We use Advanced eDiscovery analytics to thread email conversations, remove duplicates, identify near-duplicates, and derive themes. In some cases, we use the Advanced eDiscovery Relevance feature instead of doing heavy keyword culling. And even if we’re using keyword filtering, we always use Advanced eDiscovery export because it lets us hand off content to our eDiscovery review partner in a format that doesn’t require reprocessing.

The rest of this section describes some of the ways we use Office 365 eDiscovery at Microsoft in legal matters. The process we use would be similar for other types of investigations.

Placing content on hold
At Microsoft, we place content on hold from within the cases that we’ve created in the Security & Compliance Center. We create separate holds for each case. Because the Security & Compliance Center is fully scriptable, we save time by using PowerShell automation. One script that we created automates the daily task of applying in-place preservation to the content sources of custodians who were recently put on legal hold. Of course, the same tasks can be done manually in the Security & Compliance Center.

Each matter in our dedicated legal-hold database—provided by a third-party—has a corresponding eDiscovery case in the Security & Compliance Center. We apply in-place hold to Exchange mailboxes and OneDrive for Business sites associated with the custodians for each matter. Daily, we run a script that discerns any legal hold activity that’s occurred since the previous day.

The script looks up the Exchange and OneDrive For Business locations for each custodian and confirms whether an eDiscovery case already exists in the Security & Compliance Center for the corresponding matter. If not, it creates the case, and adds our internal eDiscovery team as members. It then creates a hold in the case, and adds the custodians’ Exchange and OneDrive For Business sources to the hold. If an eDiscovery case already exists, it confirms that the custodians’ sources are associated, and if not, it adds them to the hold.

If custodians sync their desktops with OneDrive for Business, the synced desktop content is put on in-place hold, too. Custodians can continue to work with the information and aren’t necessarily aware that the content is on hold unless we tell them.

The following script example creates a case and corresponding hold per matter. It adds the corresponding sources listed on the CSV to the appropriate case hold policy. It’s simpler than the script we use because we have a complex, hybrid environment, but it can give you an idea of just how simple it can be to automate a presumably large process in Office 365.
Office 365 meets evolving eDiscovery challenges in a cloud-first world

Table 1. PowerShell script example for in-place hold (Preservation)

<table>
<thead>
<tr>
<th>MatterName</th>
<th>Alias</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matter 1</td>
<td>aaa</td>
<td><a href="https://contoso-my.sharepoint.com/personal/aaa_contoso_com">https://contoso-my.sharepoint.com/personal/aaa_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>bbb</td>
<td><a href="https://contoso-my.sharepoint.com/personal/bbb_contoso_com">https://contoso-my.sharepoint.com/personal/bbb_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>ccc</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ccc_contoso_com">https://contoso-my.sharepoint.com/personal/ccc_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>ddd</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ddd_contoso_com">https://contoso-my.sharepoint.com/personal/ddd_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>eee</td>
<td><a href="https://contoso-my.sharepoint.com/personal/eee_contoso_com">https://contoso-my.sharepoint.com/personal/eee_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>fff</td>
<td><a href="https://contoso-my.sharepoint.com/personal/fff_contoso_com">https://contoso-my.sharepoint.com/personal/fff_contoso_com</a></td>
</tr>
<tr>
<td>Matter 1</td>
<td>ggg</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ggg_contoso_com">https://contoso-my.sharepoint.com/personal/ggg_contoso_com</a></td>
</tr>
<tr>
<td>Matter 2</td>
<td>aaa</td>
<td><a href="https://contoso-my.sharepoint.com/personal/aaa_contoso_com">https://contoso-my.sharepoint.com/personal/aaa_contoso_com</a></td>
</tr>
<tr>
<td>Matter 2</td>
<td>bbb</td>
<td><a href="https://contoso-my.sharepoint.com/personal/bbb_contoso_com">https://contoso-my.sharepoint.com/personal/bbb_contoso_com</a></td>
</tr>
<tr>
<td>Matter 2</td>
<td>ddd</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ddd_contoso_com">https://contoso-my.sharepoint.com/personal/ddd_contoso_com</a></td>
</tr>
<tr>
<td>Matter 2</td>
<td>fff</td>
<td><a href="https://contoso-my.sharepoint.com/personal/fff_contoso_com">https://contoso-my.sharepoint.com/personal/fff_contoso_com</a></td>
</tr>
<tr>
<td>Matter 3</td>
<td>ccc</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ccc_contoso_com">https://contoso-my.sharepoint.com/personal/ccc_contoso_com</a></td>
</tr>
<tr>
<td>Matter 3</td>
<td>ddd</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ddd_contoso_com">https://contoso-my.sharepoint.com/personal/ddd_contoso_com</a></td>
</tr>
<tr>
<td>Matter 3</td>
<td>eee</td>
<td><a href="https://contoso-my.sharepoint.com/personal/eee_contoso_com">https://contoso-my.sharepoint.com/personal/eee_contoso_com</a></td>
</tr>
</tbody>
</table>

# Create a remote PowerShell session in EOP
$UserCredential = Get-Credential
Import-PSSession $Session

# Import the mapping csv
$data = import-csv .\CaseCustodianMapping.csv

# Specify desired Case Members
$Members = "eDiscMgr1@contoso.com","eDiscMgr2@contoso.com","eDiscMgr3@contoso.com"

# Create each case and corresponding hold policy for the appropriate sources
$matters = $data.MatterName | sort -Unique
foreach ($Matter in $matters)
{
    $Mailboxes = ($data | where {$_.MatterName -eq $Matter}).Alias
    $URLs = ($data | where {$_.MatterName -eq $Matter}).URL
    New-ComplianceCase -Name $Matter
    Update-ComplianceCaseMember -Case $Matter -Members $Members
    New-CaseHoldPolicy -Case $Matter -Name $Matter -ExchangeLocation $Mailboxes -SharePointLocation $URLs -Enabled:1
    New-CaseHoldRule -Policy $Matter -Name $Matter
}

# Remove the remote PowerShell session
Remove-PSSession $Session

For reference information that you can use to create your own PowerShell scripts, see Office 365 Security & Compliance Center cmdlets.

microsoft.com/ITShowcase

December 2016
Using Office 365 eDiscovery search to filter content

At Microsoft, when we need to find specific content to respond to discovery requests, we use eDiscovery search in the Security & Compliance Center. Typical source locations that we search are Exchange mailboxes, OneDrive for Business sites, SharePoint Online sites, on-premises SharePoint sites, file shares, and local computers. Because outside counsel is usually the most familiar with the case, they often provide a list of relevant custodians and complex queries to run against their content source locations. Our internal eDiscovery team runs the queries on their behalf to find the subsets of potentially relevant content required to respond to the discovery requests.

We’re able to search content in Office 365 right away because it’s already indexed. We go to the appropriate case in the Security & Compliance Center and run eDiscovery searches across the relevant source locations, without the preliminary step of having to collect and reprocess the content in a separate tool.

While most content is searchable in Office 365, some of it isn’t fully indexed. This subset may include image files, oversized spreadsheets, password-protected files, and rarely, unsupported file types. For every source location searched for each case, we also export unindexed items, so we can remediate items with unsearchable content. Even if most of a document is searchable, if it’s flagged with an index error, we download and reprocess it to ensure that our searches are as comprehensive as possible. With OCR coming to Advanced eDiscovery, the number of files requiring remediation is dropping sharply, but error handling and remediation remains an important eDiscovery task.

We also collect content from desktop computers to search. A paralegal collects this content in a careful manner so as not to impact native metadata, and uploads the content to Azure file shares. We import the local PST files in those collections to Exchange Online mailboxes and search them alongside the rest of our Office 365 sources. The rest of the content on the Azure file shares from those local collections is crawled and indexed by an on-premises SharePoint farm, and we use our SharePoint eDiscovery Center in that farm to search the content. In the near future, we’ll be importing content collected from desktop computers to SharePoint Online sites and searching these using the Security & Compliance Center Content Search capabilities.

Before Office 365 eDiscovery was available, we used to manually collect content from the various sources. It took significant time to gather the large volume of content and load it into an offline processing tool. Then, after the content was aggregated, we had to reprocess it on comparatively slow offline appliances. With all of the collection, processing, and remediation required, it could take two to three weeks before we could give outside counsel the documents they requested. Today, we do most of this work in hours, not days or weeks. We start to export content on the fly and have it ready for counsel to load into a review tool by the end of the day.

With Advanced eDiscovery, we’re also able to deliver content to reviewers in a format that virtually eliminates the need for additional processing on their end. They can load the content directly into a review tool and use the duplicate, near-duplicate, and threading information provided by Advanced eDiscovery. This means we don’t incur additional time and expense for reprocessing.

We’ve been able to do more of our discovery processes in-house using the eDiscovery features built into Office 365. As a result, by reducing the amount of manual work required for discovery requests, we’re saving about $4.5 million annually.
The following figure shows a typical case at Microsoft, where we’ve used Office 365 eDiscovery to reduce the amount of content that we send to our attorneys for legal review. Of 50 people typically placed under hold in an average case, we generally search the content of 13. Using Office 365 eDiscovery search, we typically reduce the average review set from 644 GB to 24 GB. With a 30 percent deduplication reduction, and another 26 percent from email threading, Advanced eDiscovery further reduces the review volume to 13.5 GB.

![Diagram](image.png)

*Figure 6. Average reduction in our per-case review volume in 2016 from using Office 365 eDiscovery capabilities*

**Automating search and export**

At Microsoft, we frequently use a PowerShell script to run searches and export files. Although we could use the Security & Compliance Center user interface for searches, we prefer to automate the process and work in bulk. Using offline CSV files, we specify custodian aliases, source location URLs, and the search queries to run against them. This way, we can create and run many searches in a matter of seconds. Rather than manually adding one source at a time to a given search, we add as many sources as we need to, all at once. The script runs the searches and queues exports of the results.
The following table gives an example of the type of script we use to do this.

**Table 2. PowerShell script example for search and export**

<table>
<thead>
<tr>
<th>Alias</th>
<th>URL</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaa</td>
<td><a href="https://contoso-my.sharepoint.com/personal/aaa_contoso_com">https://contoso-my.sharepoint.com/personal/aaa_contoso_com</a></td>
<td>from:johndoe AND (vacation OR OOF OR OOO)</td>
</tr>
<tr>
<td>bbb</td>
<td><a href="https://contoso-my.sharepoint.com/personal/bbb_contoso_com">https://contoso-my.sharepoint.com/personal/bbb_contoso_com</a></td>
<td>ediscovery AND tar</td>
</tr>
<tr>
<td>ccc</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ccc_contoso_com">https://contoso-my.sharepoint.com/personal/ccc_contoso_com</a></td>
<td>((aaa OR bbb) NEAR(25) (ccc OR ddd OR eee)) NOT (x OR y OR z)</td>
</tr>
<tr>
<td>ddd</td>
<td><a href="https://contoso-my.sharepoint.com/personal/ddd_contoso_com">https://contoso-my.sharepoint.com/personal/ddd_contoso_com</a></td>
<td></td>
</tr>
<tr>
<td>eee</td>
<td><a href="https://contoso-my.sharepoint.com/personal/eee_contoso_com">https://contoso-my.sharepoint.com/personal/eee_contoso_com</a></td>
<td></td>
</tr>
</tbody>
</table>

```powershell
# Create a remote PowerShell session in EOP
$userCredential = Get-Credential

# Import the search csv
$data = Import-Csv .\SearchTest.csv

# Define the name of the case and the search
$Case = "Test Case"
$SearchName = "Test Search 1"

# Concatenate queries then create the search, add the sources, and run the search
$mailboxes = ($data | where {$_.Alias.length -gt 0}).Alias
$urls = ($data | where {$_.URL.length -gt 0}).URL
$queries = ($data | where {$_.Query.length -gt 0}).Query
Foreach ($1 in $queries) {
    $query += $1 + '(c:s)' |
} $query = $query.TrimEnd('(c:s)')

eoNew-ComplianceSearch -Case $case -Name "$Case $SearchName" -ExchangeLocation $mailboxes -SharePointLocation $urls -ContentMatchQuery $query
Start-ComplianceSearch -identity "$Case $SearchName"

# Start the export
New-ComplianceSearchAction -SearchName "$Case $SearchName" -export -IncludeSharePointDocumentVersions $true -Format FxStream

# Remove the remote PowerShell session
Remove-PSSession $Session
```
Validating search results

For legal and regulatory cases at Microsoft, we work collaboratively with our outside counsel to validate and refine the search results. We aim to provide the right items and eliminate irrelevant content and false positives. Having interviewed the witnesses and search custodians, outside counsel has the best understanding of how the custodians work, including codenames and jargon, and they use that familiarity to conceive lists of complex search terms to use in our queries.

Office 365 makes this collaboration easy. After running the initial searches, we view keyword statistics to evaluate query performance. With outside counsel, we review the number of items returned by each query, focusing specifically on those that are either too broad or too narrow. Often through Skype desktop sharing sessions, we collaborate on query revisions to achieve a more appropriate cull, ensuring that we find the content we need while reducing obviously non-responsive material. This typically reduces the volume of the raw content by 96 percent.

Using Advanced eDiscovery at Microsoft

Instead of immediately downloading search results from eDiscovery, we’ve started using Advanced eDiscovery for further processing and analytics. It saves time and expense by eliminating the need to process the files so the review tools can ingest them. It also streamlines the review by batching like documents and only the unique and inclusive items from email threads for us to give to the same reviewer. This dramatically increases throughput and accuracy throughout the review process.

At Microsoft, image files are fairly common. People paste screenshots into messages and documents, scan hardcopy items to TIF or PDF files, and receive faxes as image-only email attachments. Exchange and SharePoint index each document’s extractable text, but not the words shown in the image files. The new OCR feature in Advanced eDiscovery gives the machine-learning process an even deeper understanding of the documents by including the text extracted from image files, too.

Exports from Advanced eDiscovery automatically include items that couldn’t be fully processed, segregated from the rest of the content in an “Error_Files” directory. Because Advanced eDiscovery reindexes the data more deeply—removing file size limitations, supporting more file types, and adding OCR support—the volume of index errors in Advanced eDiscovery is significantly smaller than with Office 365 eDiscovery search. This makes it affordable to have a review vendor remediate that subset for us, rather than doing the remediation ourselves.

We also give outside counsel access to cases in Advanced eDiscovery so they can train the system using Relevance for predictive coding. To help counsel more quickly decide whether a document is relevant, we use persistent keyword highlighting to make words in a document that relate to a specific topic apparent at a glance. When Advanced eDiscovery displays a document for tagging during the Relevance process, it highlights the words that we entered. Counsel can skim through a document and immediately spot the portions of it that apply to the issue at hand.

When using Relevance for predictive coding, if the training is to be done by a team of individuals, we ask them to sit together to collaboratively train the first assessment set of documents. This prompts them to discuss and resolve ambiguity across their perspectives, and ensures a higher degree of consistency in their subsequent training rounds that they do individually.

After the system is stabilized and we achieve the appropriate Precision and Recall level for the individual matter, we test the results, reviewing a statistically representative sample of items that, according to the algorithm, isn’t likely to be responsive, to validate accuracy. If the responsive content in that test set is higher than an acceptable level, we do additional training to capture more of the relevant content.
Managing small-scale internal review projects

Whether it’s for early case assessment, responding to a subpoena, or conducting a quick review of several custodians’ relevant files for an internal investigation, we often forego the effort and expense of loading search results into a full-fledged review tool. Instead, our internal team parses through exported Office 365 content on their local computers. The Export_List.xlsx spreadsheet that’s exported from Advanced eDiscovery helps facilitate such projects beautifully. In these instances, using email threading and eliminating duplicates, we’re able to export just the unique, inclusive items to give the reviewer the ability to see the full picture in the fewest files possible.

The spreadsheet lists each item in a table that’s convenient to sort and filter. Each item has an embedded hyperlink so the reviewer can open and review the downloaded copies with a single click. They can make annotations and comments on the items directly in the spreadsheet without compromising native file metadata because the original values were already logged. Our reviewers often add columns to track responsiveness, privilege, sensitivity, and so forth, so that when they’re finished, they can easily filter the list to show the subset of items that they may ultimately wish to share with others, or even produce, whatever the case may be.

Our key takeaways for eDiscovery

Some key takeaways from our experiences working with Office 365 eDiscovery are:

- Office 365 eDiscovery dramatically reduces costs and increases efficiency. It lets us seamlessly enforce legal hold compliance for a significant majority of our custodians’ content.
- Advanced eDiscovery significantly reduces our review volumes and costs. All content that requires review or production benefits from the advanced analytics—like grouping email threads, near-duplicates, and themes—even when we don’t use predictive coding.
- Predictive coding works best when we have 10,000 or more items to review. The other analytics features work well with any collection.
- We check unsearchable items and perform a reasonable amount of remediation to find potentially relevant content.
- PowerShell automates complex and repetitious tasks, so we save even more time by taking advantage of the scriptable interface of the Security & Compliance Center.
- Consistency, repeatability, and objectivity are key to defensibility in eDiscovery.
- eDiscovery partners can help your organization establish objective eDiscovery processes.

For more information

Microsoft IT Showcase
microsoft.com/ITShowcase

Auditing and Reporting in Office 365
eDiscovery in Office 365
Manage eDiscovery cases in the Office 365 Security & Compliance Center
Office 365 Advanced eDiscovery
Office 365 Security & Compliance Center cmdlets

© 2016 Microsoft Corporation. All rights reserved. Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners. This document is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY.