INFORMATION PAPER

Advanced ECA Techniques in Microsoft Purview eDiscovery (Premium)

A Process Design Framework from HaystackID®





Introduction

Early Case Assessment (ECA)¹ is a set of techniques used in the field of eDiscovery to help rapidly compile and examine data related to audits, investigations, and litigation. While not a specific stage in the EDRM[™] (Electronic Discovery Reference Model) process, ECA is regularly associated with the processing and analysis of electronically stored information (ESI) and can provide striking volume reduction and evidence identification benefits when properly applied to eDiscoverycentric projects. Despite its potential power as a discovery tool, ECA is sometimes underused and neglected due to intricacies associated with the challenge of employing multiple technologies and tools to accomplish specific ECA objectives. In this paper, Microsoft Purview eDiscovery (Premium)² is highlighted as an advanced ECA tool that may, based on its integrated and unified capabilities, be able to remove the obstacle of real and perceived complexities of ECA for eDiscovery professionals and help them confidently leverage the power of ECA in their cases and matters.

Microsoft Purview eDiscovery (Premium) and Microsoft 365 are registered trademarks of Microsoft Corporation in the United States and/or other countries. All other product and company names mentioned herein may be trademarks of their respective owners.

Early Case Assessment

Despite its longstanding presence in the eDiscovery lexicon, ECA is not an official EDRM stage. Technically spanning multiple stages, ECA is most commonly associated with Processing and Analysis. ECA is also a technique that can be leveraged to achieve some of the most transformative results in an eDiscovery project – particularly in terms of volume reduction and expedited identification of high-value content. Unfortunately, ECA is generally not as well-understood as other eDiscovery processes, and with widely varying ECA capabilities among different technologies, it is sometimes avoided altogether. As this decision sometimes proves highly detrimental in terms of increased project time, effort and cost, it is advisable for case teams to fully consider the opportunities available via a well-executed ECA workflow and avoid missing out on its many potential benefits.

Although eluding any universally-recognized definition, ECA³ refers to "...the process by which organizations rapidly gather and analyze data about potential matters in eDiscovery, compliance, or internal investigations to reach informed decisions about how to proceed..." Another definition describes "the process of attempting to quickly surface key electronically stored information (ESI), paper documents, and other potential evidence early on in a legal matter." ECA combines legal and technological analysis to quickly gather important case information, reduce the amount of data (and cost), and find relevant evidence as soon as possible in new matters to assist stakeholders in future decision-making.

Advanced ECA in Microsoft Purview eDiscovery (Premium)

With data culling (i.e., the rapid reduction of data volume via the identification and suppression of demonstrably irrelevant content, performed in a legally defensible manner) serving as a primary objective of an effective ECA workflow, the capabilities of eDiscovery (Premium) are of central importance.

The current state form of eDiscovery (Premium) provides organizations with a sophisticated toolset that eclipses those historically available outside of specialized, downstream eDiscovery platforms.

In a manner more on par with industry-leading applications, eDiscovery (Premium)'s native feature set includes: advanced indexing, advanced search capabilities, powerful filtering options, data reporting and visualization features, email threading, email domain extraction, textual near-duplicate detection, document theme extraction, data analytics, and predictive coding.

Advanced ECA Workflow Components

Analytics Processing

 Analytics processing in eDiscovery (Premium) is performed as a secondary (and optional) post-processing operation, manually executed on a "review set," to which a selection of documents has been committed for review. This process is straightforward to perform, with a few additional settings to be configured beforehand.

eDiscovery (Premium) > Cases >	s Hold Processing Exports Jobs Settings
Saved filter queries 🗸 🗟 Save the query 🙀 Reset 5	Search & analytics
Keywords: Any 🗠	Configure how search will work within your organization.
Overview Image: Analytics Image: Actions ✓ Tag Image: Imag	 Duplication identification & grouping Near duplicates/email threading Themes Email threading OCR settings
 ☑ ☑ ☑ △ △ ☑ △ △ △	Texts to ignore Select
Accepted: Digital O O Ready A	
Figure 1	Figure 2

- When analytics processing is complete, a new "Saved filter query" is automatically created, and will appear as "[AutoGen] For Review." Note the pre-filtering item count of 5166.
- When this filter is selected, eDiscovery (Premium) uses the new analytics results to suppress content⁴ like non-inclusive email items, near-duplicate email items, duplicate documents, etc.
 This technology reduced the very small collection of only 5166 documents by 63%, down to only 1880.





Filtering and Analysis

Like the analytics results filter, eDiscovery (Premium) includes another automaticallygenerated filter to help suppress the appearance of "Potentially Immaterial Items"
a common problem in downstream document review.

[AutoGen] For Review Film. × [AutoGen] Potentially Immaterial Items Immaterial Items Filter Test (08.18.22) Immaterial Test (08.18.22) Immaterial Subject/Title Status Date Sender/Author		nite	r queries 🗸 🔚	Save the query	🐺 Reset 🦷 Filt	ers	
Filter Test (08.18.22)		[A	utoGen] For Review	v	Fil ×		
Pitter lest (00.10.22)	J	[A	utoGen] Potentially	Immaterial Item	15		
Subject/Title Status Date Sender/Author File clar		Filt	ter Test (08.18.22)		Tag files	Manage 🗸 🜔 Ref	resh
		C.	Subject/Title	Status	Date	Sender/Author	File class

Figure 5

- Another of eDiscovery (Premium)'s important but underutilized analysis features is its easily-customizable document grid or table view. The table view supports a high-level evaluation of the review set content and a much more focused analysis.
- This view is easily configured from its default state via the "Customize columns" button to display whatever available metadata best supports the user's current objective. An example of this might involve a hypothetical effort to determine whether the subject of an ethics investigation communicated with a third party in a manner that violated the organization's policy on such interactions.

		5166 items	[≣ Group ∨	Customize co	lumns
File class	Bcc	Ce		Recipients	Cust
Document					lidial
Email				Lynne <ly< td=""><td>ionis</td></ly<>	ionis

Figure 6

Choose the columns to displa	ay or drag and
drop to reorder. Or press Cor	1
rrow keys to move the item	
one position at a time to achi eordering and press arrow k	
he checkbox.	.,
Search columns	
Search	Q
File type icon	
File type icon	
Subject/Title	Ц
Status	ü
	- 3
S Date	Ξ
Sender/Author	H
File class	-11
Sec.	-H
V Cc	
Recipients	E

 Although revised in December 2022 to support a paginated format similar to the longstanding user experience in platforms like Relativity[®], the document grid also remains available as a virtual table. This means that, for result sets that exceed the available screen real estate, the view will automatically update with additional documents when scrolling, without the need to navigate to subsequent pages manually.

🔓 Manage

🖸 Load sets

👶 Non-Office 365 data

Search profile view

🔿 Tags

C Refresh

This is a much-appreciated feature when applying the table view to perform iterative ECA tasks and can be enabled as follows

 The following collection of screenshots reflects the complete list of filters currently available in eDiscovery (Premium):

Figure 9		
∽ Search (2)	Item properties (60) Annotations Author Bcc Cc Comments	Is from document version Is inline attachment Is modern attachment Item class Last modified Location name Meeting end date Meeting start date
KQL	Compliance labels	Message kind
Keywords Analytics & predictive coding (9) Attorney-client privilege score Dominant theme Inclusive type Marked as Pivot Marked as representative NdEtSortExclAttach Pivot ID Potentially privileged Themes list	Compound Path Contains Deleted Message Contains Edited Message Conversation Topic Conversation Type Created Time Custodian Data source Date Document Company Cocument Subject Email internet headers Email internet headers	Native file extension Native file name Participant domains Participants Processing status Received Recipient domains Recipients Sender Sender domain Sender/Author
 ○ ID's (8) ○ Conversation ID ○ Email set ID ○ Family ID ○ File Id ○ Group Id 	Email security Email sensitivity File class Has attachment Has attorney participant Ignored processing errors Importance	Sent Size Subject Subject/Title Tags Teams Channel
Immutable ID		Title Was remediated

Title

Unique in email set

To

Is email attachment

Is from Pst

Is embedded document



Load Id

Representative ID

Word count

Work product



The opportunity for legal teams to directly leverage highly

accessible features like the table view as early as possible to rapidly develop new insight represents a profound shift for corporate law departments.

The efficiency in simultaneously leveraging institutional knowledge and legal analysis across a broad range of real-world project scenarios while maintaining all data within the organization's Microsoft 365 (M365) tenant security environment represents a win for both legal and information • Zooming into a specific use case for these features, with the "Message kind" filter selected within a review set, the following dialogue appears, with available selections for any combination of email, meetings, tasks, contacts, Microsoft Teams, and Microsoft Yammer:

aved filter queries \vee	El Save the	ouerv 😒	Reset S Filters
Message kind: Any	-	arci -	
viessage kind: Any			
Equals any of		\sim	🐼 Tag files 🛛 Mar
			Date
email	1.86 K	-	0.1.21.2012.7/
meetings	308		Oct 24, 2012 7:
tasks	122		Aug 15, 2022 9:
contacts	78		Aug 17, 2022 2:
microsoftteam	ns , im 38		Aug 17, 2022 2:
yammer	24	-	Aug 5, 2022 4:3
			Aug 5, 2022 4:3

Figure 11

- Additional logic options are also provided to more precisely filter the specific content of interest:
- Building on this concept, additional filters can be simultaneously added in order to apply multiple criteria.⁵ This feature allows users to search through a complex dataset for specific information, similar to what has typically only been possible after data has been processed and indexed in a downstream eDiscovery platform. The benefit of this feature is that you can perform these targeted searches before any data has left the organization's M365 tenant.
- The screenshot below depicts a more advanced, multi-criteria scenario, with a keyword search for meet* OR review* OR schedule, within Exchange data only, and a date range of 8/1/2022 to 8/5/2022.

eDisco	overy (Premium)	> Cases > i	ina XiA	
Saved	filter queries 🗸	🖉 🗟 Save the	e query 🕤	🥇 Res
Messa	age kind: Any	\sim		
E	quals any of		~	a
E	quals any of			
E	quals none of			н
C	Contains all of			
E	tasks	122		
	contacts	78		
	microsofttean	ns , im38		
E	yammer	24	-	
		an ooptor ne rien	. neouj	_





Email Domain Filtering

- Another capability introduced via eDiscovery (Premium)'s analytics processing is the extraction of email domain information and the resulting ability to isolate, then exclude, or include content by domain.
- Returning to our prior hypothetical of an ethics investigation, consider a set of email search results with a much higher-than-expected volume based on what is already known to be a minimal number of possible results. In such a situation, an assessment of email domains⁶ could provide needed insight – perhaps in the form of a problematic, false-positive search hit associated with a specific domain.

In such a case, let's imagine that filtering by sender domain immediately reveals obvious, falsepositive participant domain hits for: amazon.com, linkedin.com, and tradejournal.com – all of which can't possibly be within the scope of the investigation.

In this situation (in addition to the broader insight provided by easily viewing the complete list of sender domains), applying a simple "Equals none of" filter on those three domains would reduce the document count by over 4500 items in just a few clicks.

quals any of		~
î↓	î↓	
amazon.com m365x	ft.com 3.69 K	
linkedin.com enron.com	695	- 1
tradejournal.com	133	- 1
] yammer.com	128	- 1
] planner.office365.com	111	- 1
] enron.net	45	- 1
juno.com	38	- 1



Keyword Search

- eDiscovery (Premium)'s keyword search capabilities⁷ provide flexibility to be applied, either separately or in combination with its array of filters, to achieve precise control over the scope of a review set. It supports efficient, iterative fine-tuning of search parameters to help legal teams better identify content worthy of scrutiny by human reviewers.
- In addition to eDiscovery (Premium)'s primary keyword search capability, even more extensive capabilities are available via the KQL-based "Advanced Query Builder." KQL⁸ stands for "keyword query language" and is designed to retrieve data based on the combination of keywords and conditions specified according to the KQL syntax.
- Microsoft updated eDiscovery (Premium)'s Advanced Query Builder in 2022 for an improved user experience in constructing complex queries. It allows users to "add conditions or add condition groups that are made up of multiple conditions that are logically connected by AND OR relationships."9

∿ Keywords	
Equals any of	
patent.apreement.waiyer.copyright	
AND ~	
∑ Date	
Between	
2022-04-27 🔲 2022-05-06 📼	
AND ~	
> Participants	
Contains all of	
chrisi, brucea, lisam	
chrisi brucea lisami	







Reporting

- Multiple reporting options are available in eDiscovery (Premium), which facilitate the assessment of review set content and further support ECA efforts.
- The following are two different examples of the types of fixed reporting provided in the review sets tab:

Summary	
This review set has 1 load set(s). A load set or ingested into a review set from non Off	is the collection of documents added into a review set at a single time from Office 365 ice 365 sources.
Total extracted documents:	5166
Total processed items:	4436
Total number of failed extracted documents:	20

Figure 16

Target Population	Documents (excluding attachments)	Emails	Attachments
No statistics available	5 23 250 750 150 UNGUE.NE. DIACT.DUP	0 % 20 Bb 40 5 REST REST RELUSIVE_ PACILISIVE	0- 0 80 50 80 200 8 0 DURATIACI () UNCULATI.
Number of documents by file type No statistics available	Documents by source No statistics available	Documents aggregated by process Prostet/Dochenter legges Documents Kontect/Documents	-
		5 S B	0 w h h

Figure 17

Search Profile View¹⁰

• A final aspect of eDiscovery (Premium)'s support for advanced ECA activities resides in a corner of the Review tab UI that is very easily overlooked. However, the visual analytics capabilities that lurk there are easily the most impactful of its feature set.

Combining aspects of eDiscovery (Premium)'s analytics processing and search capabilities, the "Search profile view" or Review tab "dashboard" supports dynamic visualizations based on a selection of available pivot fields, as well as filter state.

What this functionality provides is the ability to create visualization widgets, which represent the current result set via one of four available formats:

- 1. Pie/ring chart
- 2. Horizontal bar graph
- 3. Vertical bar graph/chart
- 4. Line graph

By default, three widgets are included:

- 1. Pie/ring chart on ItemClass
- 2. Bar chart on RecipientDomains
- 3. Pie/ring chart on the To field





• The following example illustrates several additional widgets, reporting results based on a limited number of extensions via the native file extension filter:



Figure 19

In the following example, the visualizations are modified by applying two additional filters

 a keyword query and a date range. The visualization widgets update to reflect the new
filters and report on the new totals for each category:

Search profile view					
Save the query Date: 8/1/2022-8/5/2022 Keywords Keywords	s: meet* OR review* OR schedul* 🚿 Native file ext	ension: msg.docculsk.pptu/html.txt.csk.pdf.one 😤			X Clean all 🐺 Filter
→ ħæw.widget ×					
ItemClass Pie Chart	RecipientDomains Bar Chart	File Extension File Chart	Custodian column duart.	File class pie chart	Hile Class Column Chart
			800		80
	-00	ng	400		***
			200		
	1	bind 26	,		
UMAtate Smale	1191 4111	ar mais 💼 teleni.	4111111	📕 Emai 📕 Conversation 🔳 Att	adamen d
Thepes.		1940			
united male L.	412				
people a					
📕 united male temale states trivee 📒 oed	pa.				

Figure 20

• The following are examples of the individual widgets, which are viewable in higher resolution form via:



















Conclusion

Although not a comprehensive survey of the capabilities available in eDiscovery (Premium) to support ECA workflows, the information presented should be construed as a broader framework for process design, and provide food for thought from which more specific, customized workflows can be developed.

References:

¹ Also referred to Early Data Analysis or EDA.

² <u>https://learn.microsoft.com/en-us/microsoft-365/compliance/overview-ediscovery-20?view=o365-worldwide.</u> Microsoft Purview eDiscovery (Premium) is also commonly referred to as "Premium eDiscovery," "Premium" or "eDP."

³ See the following links for a more in-depth discussion of ECA: <u>https://zapproved.com/blog/eca-early-case-assessment/.</u> <u>https://www.exterro.com/basics-of-e-discovery/early-case-assessment.</u>

⁴ See the following Microsoft article for more information on this functionality: <u>https://learn.microsoft.com/en-us/</u> microsoft-365/compliance/analyzing-data-in-review-set?view=o365-worldwide#using-the-for-review-filter-query.

⁵ It's important to note that, by default, filters are evaluated in eDiscovery (Premium) with Boolean AND logic applied between them when multiple filters are used, which best supports the objective of filtering content. In order to perform the most flexible and advanced searches possible in eDiscovery (Premium), the KQL (keyword query language filter) should be used, which supports Boolean OR logic, among many other parameters, and is addressed in the next section.

⁶ Sender, recipient, and participant domains are all supported.

⁷ https://learn.microsoft.com/en-us/microsoft-365/compliance/keyword-queries-and-search-conditions?view=o365-worldwide.

⁸ <u>https://learn.microsoft.com/en-us/microsoft-365/compliance/ediscovery-kql-editor?view=o365-worldwide.</u>

⁹ <u>https://learn.microsoft.com/en-us/microsoft-365/compliance/review-set-search?view=o365-worldwide#advanced-query-builder</u>.

¹⁰ https://learn.microsoft.com/en-us/microsoft-365/compliance/advanced-ediscovery-dashboard?view=o365-worldwide.

Learn More. Today.

<u>Contact us today</u> to learn more about our Microsoft 365 capabilities and how we can help assess, augment, accelerate, and support your cyber, data, and legal discovery operations.

About HaystackID[®]

HaystackID solves complex data challenges related to legal, compliance, regulatory, and cyber events. Core offerings include Global Advisory, Data Discovery Intelligence, HaystackID Core[®] Platform, and Al-enhanced Global Managed Review powered by its proprietary platform, ReviewRight[®]. Repeatedly recognized as one of the world's most trusted legal industry providers by prestigious publishers such as Chambers, Gartner, IDC, and Legaltech News, HaystackID implements innovative cyber discovery, enterprise solutions, and legal and compliance offerings to leading companies and legal practices around the world. HaystackID offers highly curated and customized offerings while prioritizing security, privacy, and integrity. For more information about how HaystackID can help solve unique legal enterprise needs, please visit HaystackID.com.