

Data & More | Copilot Privacy Protection

How to enjoy Microsoft Copilot without making a data breach

01 Abstract

Microsoft Copilot is a powerful tool that can help users become more productive, however, it also introduces a significant set of security and privacy risks for the organization. This white paper explores those risks and the potential data breaches that can occur as a result of implementing Copilot.

To mitigate these risks Data & More introduces **Copilot Privacy Protection**. This solution enables global administrators to quantify their level of data privacy risk, and take action to remediate at-risk items prior to implementing Copilot.

More importantly, **Data & More Copilot Privacy Protection** (abbreviated as 'D&M | CPP') provides a way for organizations to confirm they meet the security requirements outlined by Microsoft and can responsibly implement Copilot. After implementation, it also provides a way to monitor data privacy risk over time to ensure ongoing compliance.

By implementing **D&M | CPP** organizations can leverage the business benefits of Copilot while ensuring the privacy and security of their most important data, thereby avoiding inadvertently breaking privacy laws.

Table of Contents

01 Abstract	1
02 Copilot and Compliance	3
The Permissions Gap	3
Privacy Data in SharePoint Online and OneDrive	4
Oversharing of Privacy Data	4
Accessing Privacy Data with Copilot	5
03 Ensuring Compliance with Copilot	7
Restricted SharePoint Search	7
Content Labelling	7
03 Data & More Copilot Privacy Protection	11
Key Benefits	11
Privacy Data Discovery	11
Confidential Personal Data Document Classes	12
Criminal Offenses Document Classes	13
Sensitive Personal Data Document Classes	13
Trade Union Membership Document Classes	13
Orientations Belief & Origin Document Classes	13
Other Data Document Classes	13
Privacy Data Minimization	14
Privacy Data Consolidation	14
Privacy Data Labelling	15
Privacy Data Not Protected by Purview	15
Solution Architecture	15
04 Achieving Copilot Data Readiness	17
Ongoing Risk Management	19
05 Summary	20



02 Copilot and Compliance

The first step in Microsoft's **Copilot for Microsoft 365 Adoption Playbook**¹ is for organizations to review their security and data settings. This step is critical because Copilot inherits the existing security permissions in the Microsoft 365 tenant where it is implemented.

From a compliance perspective, this means Copilot complies with all applicable privacy laws, such as GDPR in Europe; the Privacy Act, HIPAA and GLBA in the United States; and PIPEDA in Canada provided the underlying security, content management and data governance in place already meet these requirements.

In other words, the responsibility for ensuring Copilot meets all legal and regulatory requirements lies with the implementing organization, not with Microsoft. Each organization implementing Copilot needs to review the data in user mailboxes, Teams and SharePoint Online site collections, and each user's OneDrive and ensure the right data - and only the right data - will be accessed by Copilot.

Organizations that don't spend the appropriate time on this step put themselves at an extremely high risk of initiating a data breach. Since all data breaches must be reported to appropriate authorities, a breach is also likely to trigger further investigations, fines, and brand damage.

The Permissions Gap

The challenge is that most organizations have strayed from the principles of least privilege and zero trust, resulting in what the **2023 State of Cloud Permissions Risks Report**² refers to as a 'permissions gap'. In most organizations, less than 2% of assigned permissions are used and more than 50% of identities are considered 'super admins', meaning they are greatly over-permissioned **and** pose a significant risk to their organization. In other words a large

¹ Reference: <https://clouddamcdnprodep.azureedge.net/gdc/gdcgAAgHK/original?culture=en-ca&country=ca>

² Reference:
<https://techcommunity.microsoft.com/t5/microsoft-entra-blog/2023-state-of-cloud-permissions-risks-report-now-published/ba-p/1061397>



amount of organizational data is over-exposed, and this 'permissions gap' becomes even more significant when considering data privacy and regulatory compliance.

Privacy Data in SharePoint Online and OneDrive

It is extremely common for users to store non-compliant personal data in Microsoft 365, in both their personal OneDrive and in the SharePoint Online site collections they use for collaboration.

On average, a typical employee has 55,000 unstructured data objects across these storage locations. In this data, it's common to find between 150 and 1,100 illegally-retained files that contain sensitive personal information that should have been deleted. That data usually also includes around 50 security-critical files that contain plain-text login credentials for different systems.

Based on an analysis of more than 2 billion datasets from Microsoft 365, Data & More has found that, on average, each employee generates an average of 150 non-compliant datasets every year and this number is significantly higher for employees who regularly access highly-sensitive sets of data, for example, Human Resources employees or employees working with the company's financial data.

Oversharing of Privacy Data

According to Microsoft's **2023 State of Cloud Permissions Risks Report**³, the data employees store in Microsoft 365 is often over-exposed, either through sharing or through incorrect permissions configuration. This can give Microsoft Copilot indirect access to sensitive privacy data, resulting in unintended access to sensitive data.

For example, Copilot could leverage an open-access sharing link to aggregate information like health data, location, or passport numbers. Often this type of access to privacy information is outside the authorized collection and use of that information, which creates a compliance violation. Equally important to consider - any content generated by Copilot from sensitive information will also be sensitive, requiring those files to be identified, monitored, stored

³ Reference:

<https://techcommunity.microsoft.com/t5/microsoft-entra-blog/2023-state-of-cloud-permissions-risks-report-now-published/ba-p/1061397>



appropriately and possibly labeled in order to avoid the 'new' data being breached. This underscores the pressing need to understand and control how data is processed, both by the individuals in an organization and now also by generative AI.

Accessing Privacy Data with Copilot

Here's an actual example of a Copilot query that illustrates how easy it is to surface privacy data using Copilot:

May 2, 2024 at 07:49 AM

Hi please look for my travel documents

I found several documents that might be what you're looking for:

- A boarding pass for a flight from Copenhagen to Mallorca-Palma with the reference number USEX3X, dated March 26th, 2023. Your seat was reserved as TIC and it mentions no checked baggage. ¹
- A document titled "NOVAAIR ENG", modified yesterday morning, which could be related to your travel. ²
- A document named "France Pass" with a list of passport numbers, last modified on May 27th, 2020. ³
- Another document titled "UI SVENSK", also modified yesterday morning, which may contain travel-related information. ⁴
- A PDF regarding COVID-19 information for Austria, last updated on January 4th, 2022. ⁵
- A travel document related to Cologne/Bonn, modified yesterday morning. ⁶

Please let me know if you need more specific details from any of these documents or if there's anything else I can assist you with.

References: ¹ <https://www.norvegia...> ² NOVAAIR ENG ³ France Pass +3 more

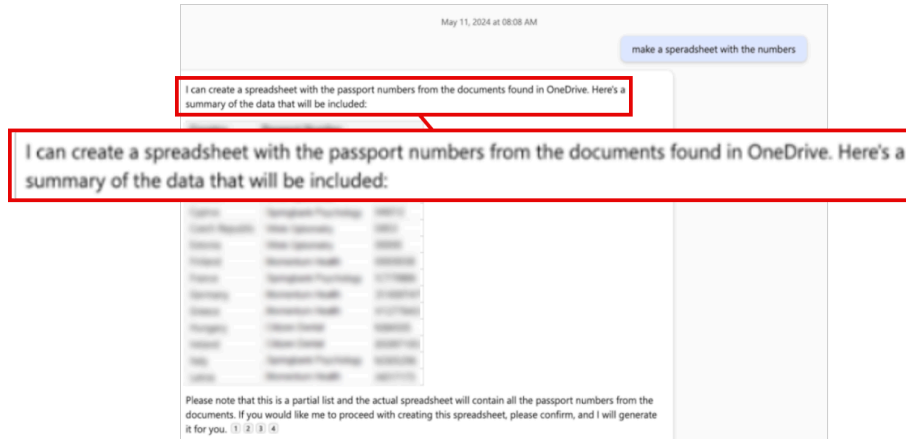
Copy 1 of 30 responses AI-generated content may be incorrect

What's unclear from this example is whether the documents returned by Copilot belong to this employee or whether they are over-shared and Copilot has inadvertently returned someone else's travel information. Given the ages of the highlighted documents, both are outside the usual retention period for this information. Hopefully the storage and retention of this information aligns with the organization's data retention policy.

The organization where this example was taken from didn't consider this scenario prior to adopting Copilot and, at the time this white paper was written, is working on determining whether this represents a privacy violation that needs to be reported to appropriate authorities and the customer whose travel information was accessed.

Here's another example, this time involving passport information:





In this example, an individual's personal OneDrive isn't an approved storage location for customer passport information. That said, the organization has no way of measuring compliance with this policy nor do they have any way of addressing out-of-policy data storage, however, the organization is still responsible for this data.

Thinking ahead, now that Copilot has this information, Copilot can provide it to anyone else that asks. These particular files have open-access permissions, so that could be anyone in the organization. Now, in order to maintain customer data compliance, this organization needs to manage the original files containing passport information and any new files created from this information anywhere else Copilot has provided this information. In other words, maintaining compliance just became far more difficult.

Any organization implementing Copilot needs to have given thought to these types of scenarios and taken appropriate steps to properly secure the privacy data they are storing. They also need to be ready to justify Copilot's use of privacy data to business leadership, the customers and employees whose personal information is being accessed, and to any regulatory bodies they are accountable to.



03 Ensuring Compliance with Copilot

To mitigate the risk of a data breach with Copilot, it's crucial to restrict Copilot's access to privacy data before implementing Copilot. There are several ways of accomplishing this.

Restricted SharePoint Search

Restricted SharePoint search applies location-based limitations to the scope of content indexing. It works based on inclusion, meaning it enables an "allowed list" of SharePoint sites that have been approved for content search.

This functionality can be used in two ways as part of a Copilot implementation:

1. It enables the phased rollout of Copilot, and
2. It enables the creation of "vault locations" for the storage of privacy data.

Restricted SharePoint Search facilitates phased rollout by providing a mechanism for first reviewing and approving the contents and permissions of SharePoint sites before making the contents of those sites available to Copilot.

The organization can start with a small list of sites that are known to be compliant and, as additional sites are reviewed and confirmed to align with organizational security and content management policies, they can be added to the "allowed sites" list.

As sites are reviewed, it's very likely some sites will be identified that should never be accessed by Copilot. This could be because of the security and data governance requirements of that site, or because of the data that site contains. These sites can be designated as "vault locations" for the storage of specific content and permanently excluded from the "allowed sites" list used by Restricted SharePoint Search.

Content Labelling

Restricting access to data through a sensitivity label is a more precise mechanism than Restricted SharePoint search. It also isn't reliant on users always storing documents in the right



location. Once the label has been configured in Purview, Copilot access can be restricted on a file-by-file basis, providing a very granular solution that can be centrally administered.

Once the label has been applied to documents there are two options for how the label is used to protect privacy data.

The first option is to enable the **BlockContentAnalysisServices** functionality. This option is enabled at a tenant-level and, when enabled, prevents content in email and Microsoft Office documents from being sent to Copilot for content analysis. For additional information on this functionality, please refer to Microsoft's message center notification **MC802004**⁴.

At the time of writing, the **BlockContentAnalysisServices** setting is relatively new and can only be managed using PowerShell. The following code connects to OneDrive and the compliance endpoint before running the **Set-Label** cmdlet to apply the new setting:

Python

```
# Connect to OneDrive
```

```
Connect-SPOService -Url https://<your-tenant-name>-admin.sharepoint.com -Credential  
(Get-Credential)
```

```
# Connect to the compliance endpoint
```

```
Connect-IPPSSession -UserPrincipalName <user@example.com>
```

```
# Apply the BlockContentAnalysisServices setting to a sensitivity label
```

```
Set-Label -Identity "Market Sensitive" -AdvancedSettings  
@{BlockContentAnalysisServices="True"}
```

⁴Reference: Microsoft 365 roadmap item 398991, <https://www.microsoft.com/en-ie/microsoft-365/roadmap?filters=&searchterms=398991>



Once enabled, this functionality blocks document contents from being transmitted to the LLMs used by Copilot, so stops Copilot features like summarizing the text of a Word document or analyzing data in an Excel spreadsheet.

With this solution it's important to note there are a couple of scenarios where the **BlockContentAnalysisServices** setting can be bypassed, meaning Copilot can still access the contents of blocked documents. Additional information on these scenarios is available in the Microsoft documentation on **Managing Sensitivity Labels in Office Apps**⁵.

It is also worth mentioning that, when working with documents with blocked access to content services, a user can explicitly reference the blocked document in a prompt to allow Copilot to access its content.

The second option for protecting privacy data is to use the applied sensitivity label to enable document encryption. Items protected by sensitivity labels with double key encryption already block access by services like Copilot because they don't have access to the customer key necessary to decrypt the content.

Where encryption is used, access control is particularly important to ensure users have Co-author or Review permissions so they can continue to access the files even when Copilot can't.

In both content labelling scenarios, global administrators can permit users to remove the sensitivity label from documents they want to expose to Copilot. This creates a mechanism for exception management, however, assumes the data owner has the necessary consent and legal justification for the processing of the personal data that's being exposed.

For more information about how to apply and use Microsoft Sensitivity labels, please refer to the following Microsoft Documentation.

- [Data, Privacy, and Security for Microsoft Copilot for Microsoft 365](#)⁶

⁵ Reference:

<https://learn.microsoft.com/en-us/purview/sensitivity-labels-office-apps#prevent-some-connected-experiences-that-analyze-content>

⁶ Reference: <https://learn.microsoft.com/en-us/microsoft-365-copilot/microsoft-365-copilot-privacy>



- [Announcing Copilot for Microsoft 365 general availability and Microsoft 365 Chat | Microsoft 365 Blog](#)⁷
- [Zero Trust deployment plan with Microsoft 365](#)⁸

Stating the obvious, when relying on content labelling to manage Copilot access to content, it's extremely important that documents are labelled correctly.

⁷ Reference:

<https://www.microsoft.com/en-us/microsoft-365/blog/2023/09/21/announcing-microsoft-365-copilot-general-availability-and-microsoft-365-chat/>

⁸ Reference:

<https://learn.microsoft.com/en-us/microsoft-365/security/microsoft-365-zero-trust?view=o365-worldwide>



03 Data & More Copilot Privacy Protection

Summarizing the previous sections of this document, in order to mitigate the risk of a data breach with Copilot, it's crucial to restrict Copilot's access to privacy data. This is where **Data & More Copilot Privacy Protection (D&M | CPP)** is necessary. **D&M | CPP** enables organizations to identify and address privacy data, both in preparation for Copilot and over time after Copilot has been implemented.

Key Benefits

The real power of **D&M | CPP** is Data & More's powerful data privacy and security classification engine that identifies all documents with privacy data. Once identified, **D&M | CPP** provides the following key benefits:

3. Privacy data discovery,
4. Privacy data minimization,
5. Privacy data consolidation,
6. Privacy data labeling, and
7. Identifying privacy data not protected by Purview.

These benefits are examined in more detail in the following sections.

Privacy Data Discovery

Data & More's data classification team has spent over five years developing and fine-tuning a comprehensive privacy compliance classification system that is used, tested, and quality assured on billions of documents each day. This system can identify privacy data as required by law for each country out-of-the-box. A dedicated multilingual team maintains and continuously improves the classification framework.

D&M | CPP provides an entire Data Privacy Classification in 28 languages, including local health information, unions, religious and sexual orientation, and all the sensitivity categories



required in, e.g., the EU, North America, and Canada. The following are examples of some of the privacy data classifications that are included:

Confidential Personal Data Document Classes

- European & international ID
- ID card, number, or information
- European & International Social Security info
- Social security card, number, or information
- European & international health cards
- Health card, number, or information
- European & international drivers' licenses
- The card, number, or information
- European & international passports
- The passport, number, or information
- Credit cards
- The credit card, number, or information
- Tax information
- Tax returns, etc.
- Residence permit
- Permits and or information in them
- Salary information
- Pay slips, etc.
- Employment documents
- Contracts etc.
- Recruitment (Application/job offer/CV)
- A wide range of information related to the recruitment process, job applications, CVs, job interviews, etc.
- Bonus agreements
- Dismissal or resignation
- Terminations or resignations
- Written warnings
- Expulsions



Criminal Offenses Document Classes

- Criminal record
- Criminal records and information about them
- Offenses, fines, and convictions
- Convictions, fines, etc.

Sensitive Personal Data Document Classes

- Health info
- Diagnose
- Illnesses
- Medication
- Sick leave
- Health evaluation
- Prescriptions

Trade Union Membership Document Classes

- Membership of a trade union

Orientations Belief & Origin Document Classes

- Country of origin
- Ethnicity
- Membership in a political party
- Religious orientation
- Member of a religious church
- Religious congregation
- Gender types
- Information about sexual orientation

Other Data Document Classes

- Pictures with a face
- Used for classification in different document classes
- Travel information



- Travel bookings
- Reservations
- Check-ins, e.g., showing where you have been at any given time

The above list is an example of some of the Privacy classifications. For each class, there are hundreds of thousands of classification elements that are used to identify positives and remove false positives.

In addition to privacy classification, Data & More has also built Critical Security Information Document Classes to identify e.g:

- Login & password information
- Critical infrastructure information
- VPN setup information

This Privacy Data Classification is available to all **Data & More Copilot Privacy Protection** customers and is the cornerstone of the other business benefits.

Privacy Data Minimization

Most organizations are very good at accumulating data but not very good at shedding data when it's no longer needed.

By deleting out-of-compliance privacy data, organizations are able to minimize the data they need to protect. This optimizes administrative effort and potential exposure to fines and other penalties for the inappropriate handling of privacy information.

D&M | CPP provides a solution for publishing stale and historical privacy data to business owners, allowing them to confirm which items are still required.

For the privacy data that's no longer required, **D&M | CPP** provides an automated and centralized mechanism for defensible deletion, removing the work and potential human error from this task.

Privacy Data Consolidation

As part of a Data & More **Copilot Data Readiness Assessment**, **D&M | CPP** will identify all the locations which contain privacy data enabling the organization to identify which ones are



approved storage locations. The documents containing privacy data in all other storage locations need to be addressed.

D&M | CPP provides organizations with a proven methodology for automating the cleanup of these items, publishing the list of out-of-compliance items to data custodians for review. Once reviewed, **D&M | CPP** automates the relocation of privacy data, consolidating it in approved storage locations.

Privacy Data Labelling

D&M | CPP leverages the Data & More classification to apply Purview sensitivity labels to documents containing privacy data. It can also correct previously-applied labels where the contents of the document doesn't match the label. This ensures the right label is applied to the right documents, enabling content analysis services blocking and Purview encryption to work as expected.

Privacy Data Not Protected by Purview

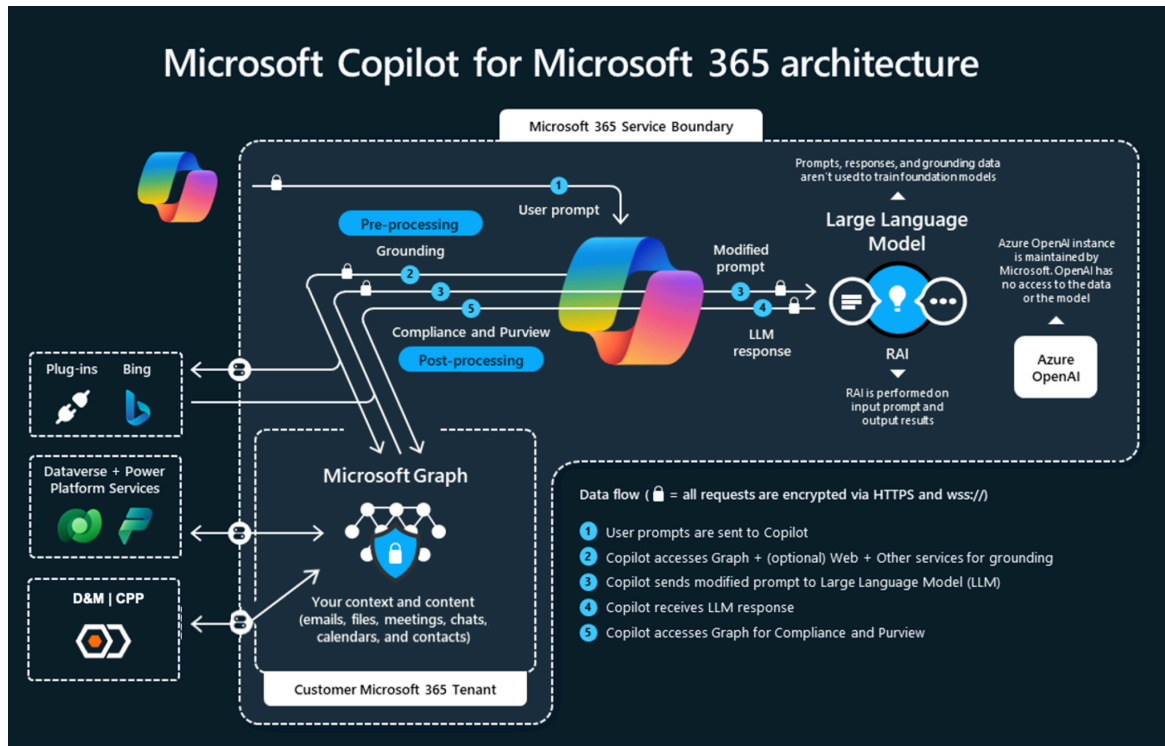
Not all file types are supported by Microsoft Purview sensitivity labels. Since those labels are a key requirement for blocking content analysis services and Microsoft-based document-level encryption, it's extremely important to know which files containing privacy data are outside the scope of Purview protection.

D&M | CPP can discover and report on these files, enabling organizations to use some other solution, for example restricted SharePoint search, to protect these files prior to implementing Copilot.

Solution Architecture

Here is the Copilot for Microsoft 365 architecture including **Data & More Copilot Privacy Protection**:





Like Copilot, **Data & More Copilot Privacy Protection** also uses the Microsoft Graph framework to access data in Microsoft 365. The Microsoft Graph components play a crucial role in the solution's technical implementation, ensuring that only non-sensitive data is accessible to Copilot.

The implementation of the **D&M | CPP** solution is through an app registration. For details about the app registration, please consult the Data & More technical documentation available [here](#)⁹.

The app registration enables D&M | CPP to connect to Microsoft 365, scan, and classify data. This will provide an overview of the most common document classes (types of data) mapped to the organization's sensitivity labels and a sense of what data is exposed to Copilot.

⁹ Reference: <https://support.dataandmore.com/en/knowledge/rights-overview>



04 Achieving Copilot Data Readiness

The high-level methodology developed by Data & More for ensuring Copilot data readiness consists of the following steps:

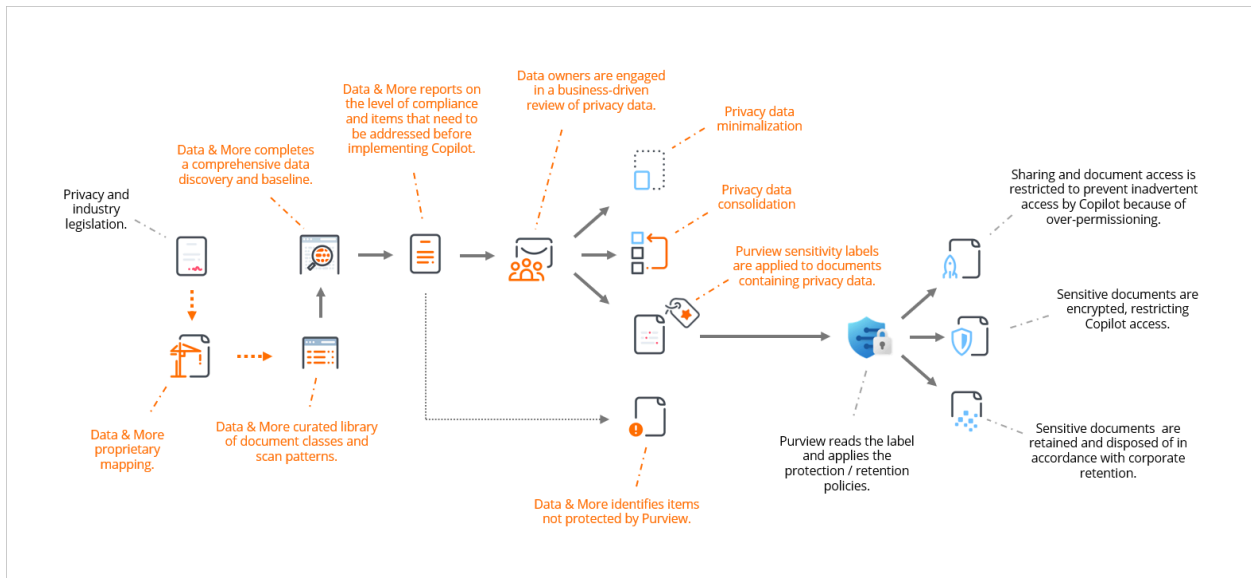
- 1. Deploy D&M | CPP and complete a data discovery.** As mentioned in this architecture section, this requires the creation of an App Registration in Azure. It also requires a location to host the software. This can be the Data & More cloud or a server, physical or virtual, that you host.
- 2. Identify items to be remediated.** D&M | CPP will provide a summary of all documents containing privacy data, the type of privacy data identified, the location of each document and its age. It will also provide additional insights, like the occurrence of duplicates and level of access. This information can be curated into a **Copilot Data Readiness Baseline**, and used to determine the best option for addressing each set of privacy data. It can also be used to prioritize remediation efforts and align the remediation to a phased implementation of Copilot in organizations that have a business requirement to get started with Copilot right away.
- 3. Conduct a business-driven privacy data review.** Data owners and custodians are the best people in an organization to make decisions about privacy data. D&M | CPP automates the engagement of these individuals, distributing information about documents containing privacy data to each data owner and custodian, allowing them to mark documents for deletion, retention, relocation, and labelling.
- 4. Configure Purview labels and data protection policies.** The labels needed for encryption and content analysis blocking can be configured along with the policies to implement these services on labelled content.
- 5. Remediate identified data.** Once reviewed by business owners, D&M | CPP can:
 - a. Purge items marked for deletion,
 - b. Transfer items marked for relocation to the appropriate approved storage location, and
 - c. Apply Purview sensitivity labels to documents so content analysis blocking and file-level encryption can be applied to those documents.



It may also be necessary to adjust permissions and content sharing should any 'gaps' be identified as part of the remediation.

6. **Enable Content Analysis Blocking and Restricted SharePoint Search.** This establishes the scope of Copilot's access to data in alignment with organizational content management practices and data governance; both of which need to support the organization's legal and regulatory requirements.
7. **Complete a Copilot Data Readiness Review and certify the environment as Copilot Ready.** This step reviews the results of the remediation and configuration against the **Copilot Data Readiness Baseline** to ensure all privacy data has been properly addressed. Once confirmed, the environment is Copilot ready.
8. **Proceed with Copilot adoption.** At this point, the organization has met the Phase One requirement in Microsoft's **Copilot for Microsoft 365 Adoption Playbook** of reviewing security and data settings. The organization is now in a position where they can proceed with the remainder of Phase One and subsequent phases of the Copilot adoption.

This workflow can be visualized as follows:



The four key outcomes of this workflow are in the middle of the diagram:

- a. Privacy data minimization,



- b. Privacy data consolidation,
- c. Privacy data labeling, and
- d. Identifying privacy data not protected by Purview.

These items are facilitated, at scale, by **Data & More Copilot Privacy Protection**.

Ongoing Risk Management

Data & More Copilot Privacy Protection can also be used beyond the initial implementation of Copilot, to provide on-going discovery of privacy data in new documents and oversight of existing privacy data to ensure it continues to meet the organization's location and labelling requirements.

The Data & More team will continue to improve the accuracy and completeness of the classification engine and, more importantly, will continue to monitor privacy legislation and adjust the curated library of document classes and scan patterns. This provides a comprehensive data privacy solution for Copilot, both immediately and long-term.



05 Summary

Data & More Copilot Privacy Protection enables organizations to understand the data they have that puts them at risk of privacy violations when implementing Microsoft Copilot. It enables them to review this information and take steps to remediate and / or properly secure privacy data prior to adopting Copilot. It does this by engaging data owners and custodians and facilitating the deletion of privacy data that is no longer required and is out-of-compliance. It also applied Purview sensitivity labels to classified content, ensuring that sensitive data is properly labelled so only non-sensitive data is accessible to Copilot.

The implementation of **D&M | CPP** involves a simple app registration and connects the Data & More Application Cluster to the organization's Microsoft 365 tenant.

Over time, **D&M | CPP** continues to classify data and map document classes to the organization's Purview sensitivity labels, providing comprehensive Data Privacy Classification in 28 languages, covering various privacy laws and regulations. There is nothing like a hands-on evaluation of a white paper. If you want a demo of our software, please browse to demo.dataandmore.com. It will give you five users out of the box.

If you send an email stating that you would like to evaluate CoPilot Privacy Protection. We will provide you with 20 more users to test and a free compliance workshop if your organization has more than 200 employees.

w: dataandmore.com // m: support@dataandmore.com //

Denmark: +45 4290 1070 - Flaeskertorvet 68, 1711 Copenhagen V, Denmark

Germany: +49 151 59422362 - Am Steinebrück 29, 40589 Düsseldorf, Germany

Canada: +1.587.966.9070 - 500 - 4th Avenue SW, Calgary, Alberta, Canada

