# Search for and find personal data

To view contributors to this article access the link below

*https://docs.microsoft.com/en-us/microsoft-365/compliance/search-for-and-find-personal-data?view=o365-worldwide*

## In this article

1. Use Content Search to find personal data
2. Search for sensitive information types using Content Search
3. Sensitive information types for EU citizen data
4. Add parameters to a sensitive information type query to hone the results

Personal data is defined very broadly under the GDPR as any data that relates to an identified or identifiable natural person that is a resident of the European Union (EU).

Article 4 – Definitions

‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

This article demonstrates how to find personal data stored in SharePoint Online and OneDrive for Business (which includes the sites for all Office 365 groups and Microsoft Teams).

Finding personal data subject to GDPR relies on using sensitive information types in Office 365. These define how the automated process recognizes specific information types such as health service numbers and credit card numbers. You can use data loss prevention policies to find personal data in mail while in transit. You can use the sensitive information types you curate for GDPR to find and protect personal information as it is sent through email. Also see Managed GDPR data subject requests with the DSR case tool in the Security & Compliance Center.

# Use Content Search to find personal data

Microsoft recommends a three-stage approach to finding personal data in Office 365. The rest of this topic provides guidance for each of these stages.

Table 1

| **Step** | **Description** |
|---|---|
| 1. Search for sensitive information types | Start by using sensitive information types to find personal data. Create a Content Search query for each sensitive information type. |

Table 1

| Step | Description |
|---|---|
| | Run the query and analyze the results.<br>If needed, add parameters to the query to reduce false positives:<br>• Count range<br>• Confidence range<br>• Other properties or operators for more complex queries<br>If necessary, modify a sensitive information type to improve accuracy for your organization:<br>• Adjust the confidence level directly in the XML.<br>• Add key words.<br>• Adjust the proximity requirements for keywords. |
| 2. Use Keyword Query Language (KQL) to find additional personal data in your environment | To find data not included in sensitive information types, use the KQL query language to develop custom queries.<br>Test the results of these searches and adjust the KQL query string until you achieve the expected result. |
| 3. Create new custom sensitive information types using the KQL queries | After optimizing KQL queries to find target data, create new custom sensitive information types using these queries. You can then use these custom sensitive information types with Content Search, in DLP policies and other tools, and within other KQL queries. |

# Search for sensitive information types using Content Search

Begin searching for personal data by using the sensitive information types that are included with Office 365. These are listed under Classification in the security center and compliance center.

This topic includes a list of some sensitive information types that apply to citizens in the European Union. Check the security center or the compliance center for new additions that can help with GDPR compliance.

Also see this article: List of sensitive information types and what each one looks for.

Sensitive information types define how the automated process recognizes specific information types such as bank account numbers, health service numbers, and credit card numbers. Sensitive information types are also referred to as conditions. A sensitive information type is defined by a pattern that can be identified by a regular expression or a function. In addition, corroborative evidence such as keywords and checksums can be used to identify a sensitive information type. Confidence level and proximity are also used in the evaluation process.

At this time sensitive information types cannot be used to find data at rest in mailboxes.

### Using Content Search with sensitive information types

Table 2

| **Step** | **More information** |
|---|---|
| Go to Content Search in the Security and Compliance Center | In the left pane of the Security & Compliance Center, click **Search & investigation** > **Content search**.<br>See Run a Content Search in the Office 365 Security & Compliance Center. |
| Create a new search item for each sensitive information type | Use the following syntax:<br>SensitiveType:"<type>"<br>For example:<br>SensitiveType:"France Passport Number"<br>Scope the search to SharePoint (includes OneDrive for Business). Make sure the syntax is exact and there are no extra spaces or typos.<br>See Form a query to find sensitive data stored on sites. |
| Review the results for each search | Look for these types of issues to determine if the query accuracy is on target:<br>• Many false positives<br>• Missing known instances of data<br>See Export Content Search results from the Office 365 Security & Compliance Center.<br>Note: if you're using Mozilla Firefox or Chrome, you might need to first download reports using Internet Explorer or Edge in order to install the required add-in. |

## Sensitive information types for EU citizen data

Start with these sensitive information types. Many more sensitive information types are coming soon for personal data in EU countries.

Belgium National Number

Credit Card Number

Croatia Identity Card Number

Croatia Personal Identification (OIB) Number

Czech National Identity Card Number

Denmark Personal Identification Number

EU Debit Card Number

Finland National ID

Finland Passport Number

France Driver's License Number

France National ID Card (CNI)

France Passport Number

France Social Security Number (INSEE)

German Driver’s License Number

Germany Identity Card Number

German Passport Number

Greece National ID Card

International Banking Account Number (IBAN)

IP Address

Ireland Personal Public Service (PPS) Number

Italy’s Driver’s License Number

Netherlands Citizen’s Service (BSN) Number

Norway Identity Number

Poland Identity Card

Poland National ID (PESEL)

Poland Passport

Portugal Citizen Card Number

Spain Social Security Number (SSN)

Sweden National ID

Sweden Passport Number

U.K. Driver's License Number

U.K. Electoral Roll Number

U.K. National Health Service Number

U.K. National Insurance Number (NINO)

U.S./U.K. Passport Number

## Add parameters to a sensitive information type query to hone the results

You can add these parameters to a sensitive information type query:

- Count range — define the number of occurrences of sensitive information a document needs to contain before it's included in the query results.
- Confidence range — the level of confidence that the detected sensitive type is actually a match, such as 85 (85%).

Syntax:

- SensitiveType:"<type>|<count range>|<confidence range>"

Examples:

- SensitiveType:"Credit Card Number|5"  (return only documents that contain exactly five credit card numbers)
- SensitiveType:"Credit Card Number|*|85.."  (confidence range is 85 percent or higher)

Note: "SensitiveType" is case sensitive, but the rest of the query is not.

You can also use properties, and operators to illustrate how you can refine your queries. For more information and examples, see Form a query to find sensitive data stored on sites.