How Data Classification Works

Wiz uses Data Classification Rules to detect potentially sensitive data, by matching your sampled data against predefined patterns. For example, a Rule that detects email addresses can include a pattern to detect `@gmail.com`. This pattern is also referred to as the Rule's Matcher logic.

When Wiz detects data that matches the pattern with high confidence, a Data Finding object is added to the Security Graph. The object is unique to the Rule and the resource it was detected on, and includes a redacted sample of the detected data.

> Only metadata such as file paths and data types, along with masked samples, are displayed in Wiz. Your data never leaves your environment. Learn about Data Finding metadata.

Data Classification Rules

Wiz includes out-of-the-box Data Classification Rules which are maintained and constantly updated by the Wiz Research Team. Once you enable data scanning for your Wiz tenant, these Rules are applied to your environment. If you'd like to identify other types of sensitive data such as your organization's unique employee ID, you can either edit the built-in Rules or add a custom Rule.

> New/updated custom Data Classification Rules take effect on the next scan, which occurs either:
>
> - On the following scan, i.e. daily for buckets, every 14 days for databases.
> - Immediately, by rescanning the Connector on demand.
>
> Updates to Data Classification Rules can add/remove the related Data Findings from the Security Graph. This is expected behavior.

Classifier types

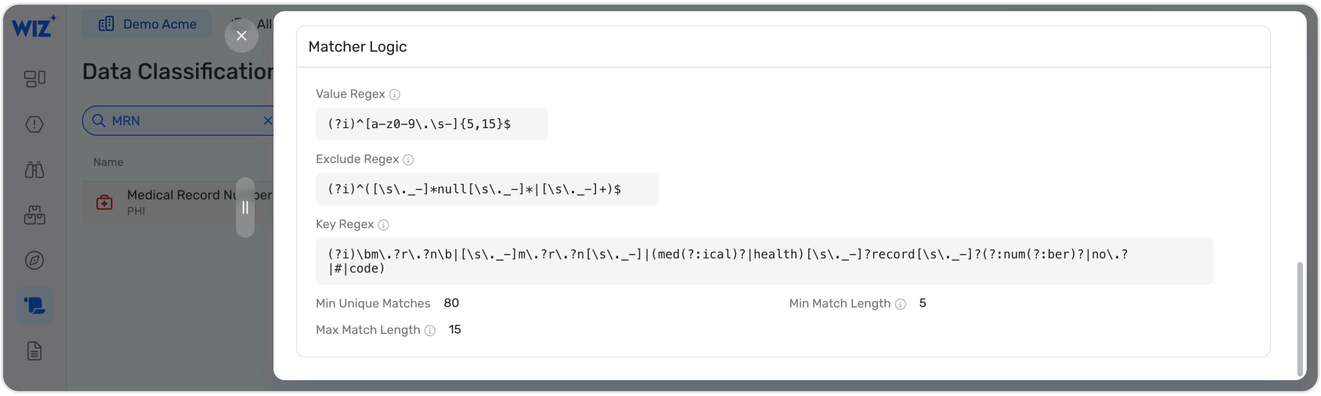
A classifier defines the pattern and/or location of various sensitive data types. For example, a credit card classifier is constructed differently from a name classifier. The Data Classification Rules then use the classifiers to search for the data, along with additional configurations and thresholds, such as the minimum unique matches required to determine a match or the expected column name, to fine-tune detections and avoid false-positives.

Below are the supported data classifier types, and how they work.

Data classifier

A data classifier detects a specific content pattern in a file, based on the data key and values. In structured files (e.g. CSV/JSON/SQL), the key will be the field/column name, and the value will be the data value (e.g. data in a table cell). The key and value patterns are defined using regex and a supporting Rego function.

A data classifier should be used when you know the exact structure of the data you are looking for, such as a Medical Record Number (MRN) or a Bank Account Number.



Matcher Logic

Value Regex

```
(?i)^[a-z0-9\.\s-]{5,15}$
```

Exclude Regex

```
(?i)^([\s\._-]*null[\s\._-]*|[\s\._-]+)$
```

Key Regex

```
(?i)\bm\.?r\.?n\b|[\s\._-]m\.?r\.?n[\s\._-]|(med(?:ical)?|health)[\s\._-]?record[\s\._-]?(?:num(?:ber)?|no\.?
|#|code)
```

Min Unique Matches 80

Min Match Length 5

Max Match Length 15

> **Coming soon!**
>
> Some of Wiz's data classifiers are in the process of being converted into keyword classifiers, such as Credit Cards, to improve our detections. These are done automatically, no action is required.

Keyword classifier

A keyword classifier defines multiple data patterns for a single data type, and should be used when you know the data type has many common variations. For example, American Express card numbers can be represented as "American Express XXX", "C num XXX", "Cards", and more. The multiple keyword variations on top of the regex pattern provide high accuracy and flexibility.

Use keyword classifiers to:

- Simplify complex key structures.
- Cover large data sets in a faster and cost-effective way.

- Detect sensitive data in both structured and unstructured files.
- Add dependency between classifiers
- Add keywords in multiple languages (Starting with English and Portugueses)

> Keyword classifiers are provided by Wiz, and you cannot create a custom keyword classifier.

See an example keyword classifier

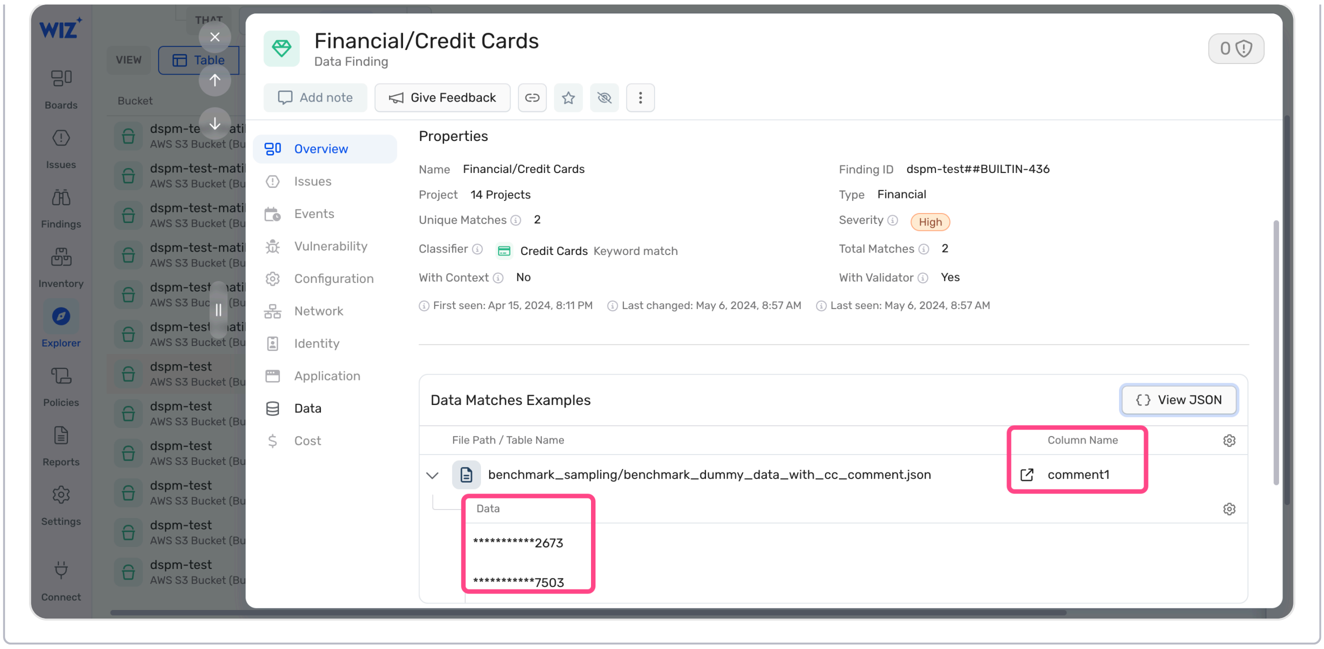
The following example demonstrates how the keyword sets of the Credit Cards classifier detect sensitive data in a JSON file that has no clear keys:

1. In the following JSON file, the keys are `sample`, `field_42`, `comment`, and `name`:

JSON

```json
[
    {
        "sample": 1,
        "field_42": "Miguel j Richardson_070-07-2282_Miguel j Richardson_Engineering geologist_Miller-Rodriguez_West Sandrachester",
        "comment1": "Regarding issue #4427 with credit card 340274359412673 not being proccesses as required, please update the client that...",
        "name": "Miguel j Richardson"
    },
    {
        "sample": 2,
        "field_42": "Timothy Williams_861-30-2155_Timothy Williams_Archivist_Nichols-Alvarado_New Jamesfurt",
        "comment1": "Regarding issue #4427 with credit card 378318627677503 not being proccesses as required, please update the client that...",
        "name": "Timothy Williams"
    }
]
```

2. Using the Credit Cards classifier, Wiz detects the Credit Card number in a field named `comment1`



Financial/Credit Cards
Data Finding

Properties

Name Financial/Credit Cards
Project 14 Projects
Unique Matches 2
Classifier Credit Cards Keyword match
With Context No
Finding ID dspm-test##BUILTIN-436
Type Financial
Severity High
Total Matches 2
With Validator Yes

First seen: Apr 15, 2024, 8:11 PM Last changed: May 6, 2024, 8:57 AM Last seen: May 6, 2024, 8:57 AM

Data Matches Examples

File Path / Table Name	Column Name
benchmark_sampling/benchmark_dummy_data_with_cc_comment.json	comment1

Data
***********2673
***********7503

Metadata classifier

A metadata classifier is used to detect specific file paths and/or file names. These should be used when you know where your sensitive data resides but you cannot define a clear content-based match for it. For example, your DMP files can potentially hold cleartext passwords or decryption keys. To detect the dump files, use a metadata classifier that searches a specific directory for `.dmp` files.

In the example below, the DMP Files - Browsers classifier uses a single Value Regex to define both the file path and file names: `(?i)(LSASS|Winlogon|credwiz)\.(hdmp|mdmp|dmp|phd|vmem|crash|core|pdb)$`.

> Metadata classifiers do not support file headers. You can define the regex to search for file path and file size only.

Data Classification Severity

Wiz assigns severities to both Data Classification Rules and the Findings they generate.

Data Classification Rule severity

A Data Classification Rule's severity indicates the impact this data type could have on your organization if it was leaked, varying from critical to low. For example, a leaked social security number is more sensitive than your company address. All built-in rules are assigned severities according to the following considerations:

Severity	Explanation	Example
Critical	A data identifier that could be used for fraud or identity theft.	- Credit card

Severity	Explanation	Example
▮▮▮▮		- Email address - SSN
High ▮▮▮▯	A data identifier that is unique to an individual and which is not publicly available, or that together with easily attainable data could be used to perform fraud.	- Bank account number - National provider identifier - Home address
Medium ▮▮▯▯	A data identifier that is not publicly available and/or that has a low impact on its own.	- Date of birth - Credit score - CVV
Low ▮▯▯▯	A data identifier that is widely publicly available and/or which is not too indicative of an individual.	- Twitter handle - IPv4 - ZIP code

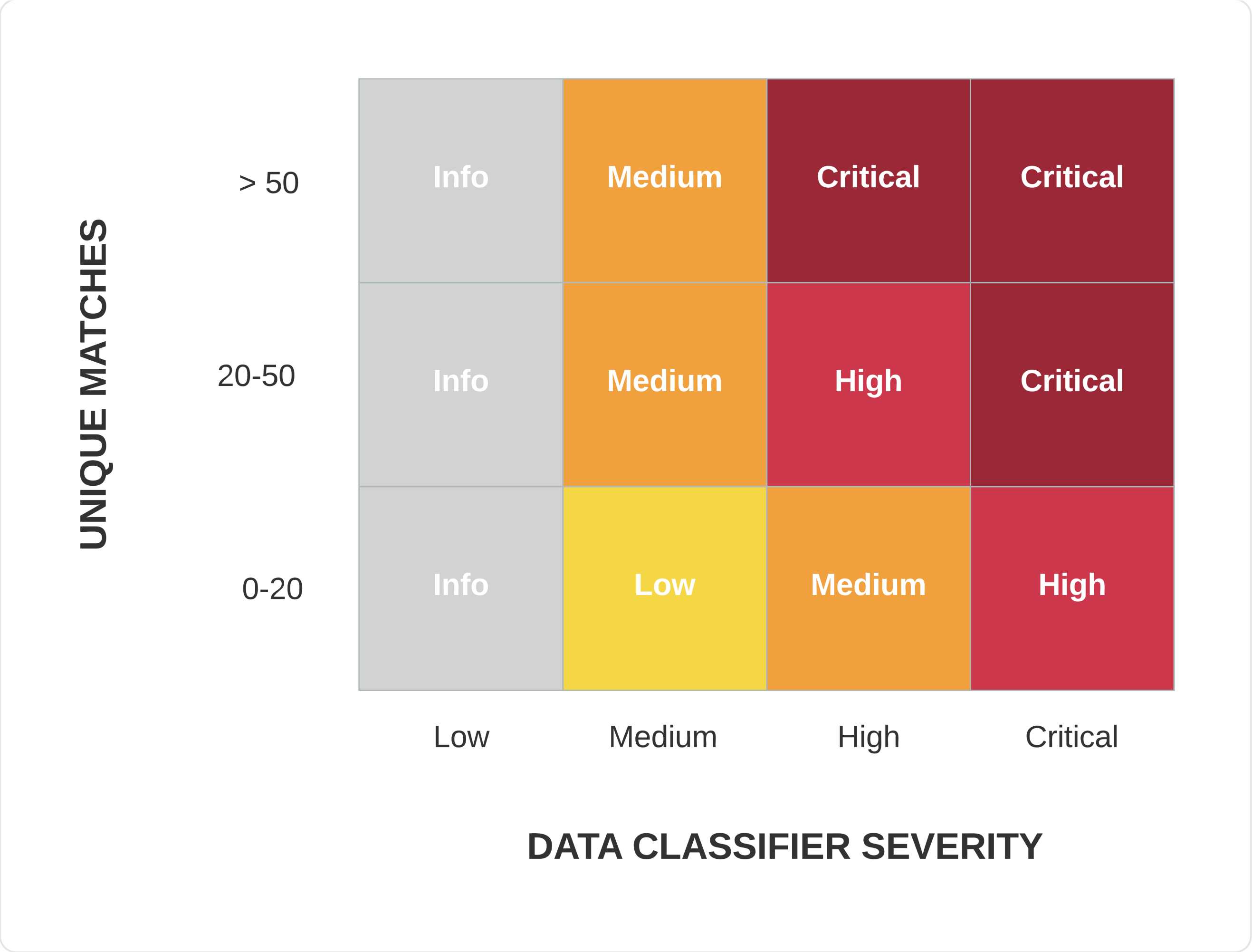
In Wiz, go to Policies > Data Classification Rules to check out the rules' severities.

Data Finding severity

Wiz assigns critical, high, medium, or low severities to Data Findings based on two factors:

- Data Classification Rule severity
- Number of unique matches within the file (e.g. a leaked file with 1000 email addresses poses a greater risk than a file with just three addresses)

The higher both of these factors are, the higher the severity. If the same Data Classification Rule applies to more than one file, Wiz will generate the finding severity according to the maximum number of unique matches.



UNIQUE MATCHES	Low	Medium	High	Critical
> 50	Info	Medium	Critical	Critical
20-50	Info	Medium	High	Critical
0-20	Info	Low	Medium	High

DATA CLASSIFIER SEVERITY

Let's look at a bucket with 2 files:

Example 1–

- File 1 has 1000 email addresses
- File 2 has 3 email addresses

Wiz will generate a single Data Finding for PII/Email in critical severity.

Example 2–

- File 1 has 100 email addresses and 100 credit scores
- File 2 has 50 emails

Wiz will generate two Data Findings: one for PII/Email in critical severity, and one for Financial/Credit score in medium severity.

FAQ

Questions? Take a look at the FAQ.

Updated 2 days ago

← Data Security Tutorials

How Data Security Works →