

ReadMeFirst for IBM StoredIQ

7.6.0.12 GDPR-Focused Basic Cartridge

The GDPR Focused Data Discovery cartridges are designed to assist organizations in their efforts to comply with GDPR, and other personal data regulations, but are not a substitute for a full organizational compliance policy.

Contents

- [Features and Benefits](#)
- [License](#)

Features and Benefits

List of supported countries

- Austria
- Belgium
- Denmark
- Finland
- France
- Germany
- Ireland
- Italy
- Netherlands
- Norway
- Poland
- Spain
- Sweden
- Switzerland
- United Kingdom

List of supported PII types

- Bank Account Numbers (International IBAN and National versions)
- National ID Numbers (sometimes multiple per country e.g tax ID, insurance, national ID card)
- Passport Numbers (currently no support for Denmark or Netherlands passport numbers)
- Phone Numbers

Caveats

The cartridge as currently configured will execute country specific patterns only when the document language is detected to match that country. For example, the cartridge will execute patterns specific to France only if the document language has been detected to be French. It uses <matchTypeFilter> statements in the concept.xml for that. This helps to prevent false positives but also means that the cartridge will not be able to find a French bank account number in an English language document or, more specifically, in a non-French language document. For example, the cartridge will not be able to find a French bank account number in most Excel spreadsheets because the detected document language will not be French, unless there is a lot of actual French text within the spreadsheet.

It is important to clearly understand the language vs. country aspect here as there is no simple 1-to-1: One language is spoken/written in more than one country and more than one language can be spoken/written in a single country. The PII rules have to be by country but the analysis system can't detect country, it can only do automatic language detection. Any filtering has to be by language (the language(s) spoken in the country for which the rule works).

Since all the language/country filtering is done via <matchTypeFilter> statements the cartridge itself is configured to claim support for "any language" (which is done by not specifying anything in the optional "languagesSupported" attribute in cartridges.properties).

Searchable ia: terms

The current cartridge.properties file defines the following six search terms:

- ia:BankAccountNumber
- ia:NationalIdNumber
- ia:PassportNumber
- ia:PhoneNumber
- ia:EmailAddress
- ia:IPAddress

The rules detect information on a finer grained level and this can be reconfigured to expose more of that detail if required for a customer use case. See below for details.

cartridge.properties contents

Cartridge Properties

name=GDPRFocusedDataDiscoveryBasic

description=SIQ cartridge to detect key personal data information ids and patterns for selected EU countries

version=1.0

Technology=RegEx

The list of supported languages is commented out since some patterns work for all

languages (e.g. IP Address) and the concept.xml has fine grained <matchTypeFilter>

```
# which can filter by language on the level of individual rules
# Removing the comment might give a marginal speed improvement but then IPAddress
# and email will no longer work on any document (only on ones detected to be in
# the list of supported languages
# languagesSupported=fr,se,en,es,no,it,de,da,nl,fi,pl

# Mapping from a supportedResult name to an UIMA type
# This file chooses a coarse grained mapping that does not expose sub-types
# E.g. BankAccountNumber has sub-types for IBAN and BBAN
# If needed these could be mapped to search results as well be adding
# additional supportedResults lines

# Main patterns
supportedResults.BankAccountNumber=GDPRFocusedDataDiscoveryBasic.BankAccountNumber
supportedResults.NationalIdNumber=GDPRFocusedDataDiscoveryBasic.NationalIdNumber
supportedResults.PassportNumber=GDPRFocusedDataDiscoveryBasic.PassportNumber
supportedResults.PhoneNumber=GDPRFocusedDataDiscoveryBasic.PhoneNumber

# Misc patterns
supportedResults.EmailAddress=GDPRFocusedDataDiscoveryBasic.EmailAddress
supportedResults.IPAddress=GDPRFocusedDataDiscoveryBasic.IPAddress
```

License

Source Components and Sample Materials

The Program may include some components in source code form ("Source Components") and other materials identified as Sample Materials. Licensee may copy and modify Source Components and Sample Materials for internal use only provided such use is within the limits of the license rights under this Agreement, provided however that Licensee may not alter or delete any copyright information or notices contained in the Source Components or Sample Materials.

IBM provides the Source Components and Sample Materials without obligation of support and "AS IS", WITH NO WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE WARRANTY OF TITLE, NON-INFRINGEMENT OR NON-INTERFERENCE AND THE IMPLIED WARRANTIES AND CONDITIONS OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

DISCLAIMER OF WARRANTIES. The included cartridge regex code is sample code created by IBM Corporation. IBM grants you a nonexclusive copyright license to use this sample code example to generate similar function tailored to your own specific needs. This sample code is not part of any standard IBM product and is provided to you solely for the purpose of assisting you in the development of your applications.

This example has not been thoroughly tested under all conditions.

IBM, therefore cannot guarantee nor may you imply reliability, serviceability, or function of these programs. The code is provided "AS IS", without warranty of any kind.

IBM shall not be liable for any damages arising out of your or any other parties use of the sample code, even if IBM has been advised of the possibility of such damages.

If you do not agree with these terms, do not use the sample code.

Licensed Materials - Property of IBM

5725-W53

Copyright IBM Corp. 2015, 2016 All Rights Reserved.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.