

# ReadMeFirst for IBM StoredIQ

## 7.6.0.12 GDPR-Focused Advanced 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

The file GDPRFocusedDataDiscoveryAdvanced.car contains the SystemT-based patterns to detect the key names of Persons, Locations/Addresses and Organizations/Companies for six languages.

- China
- France
- Germany
- Japan
- Spain
- USA

### List of supported PII types

- Person Names (names of humans)
- Locations (names of cities, countries, etc.)
- Organizations (names of companies, businesses and other organizations)
- Addresses (street addresses - strictly speaking a sub-type of Location)
- Dates (calendar dates in different, country-specific formats)
- Phone Numbers

## Caveats

Recognition quality for this cartridge is a mix between precision and recall. Recognition frequency will be much higher than the GDPRFocusedDataDiscoveryBasic cartridge, as many documents will match one of the concepts this cartridge detects. The basic cartridge contains detection logic for rare concepts like national ID numbers. The only high frequency concept in the basic cartridge is phone numbers. The advanced cartridge contains detection logic for more common concepts such as person, names, dates, etc., and will typically match most documents since most documents contain such common concepts. The intended use in StoredIQ is to create filters that use a combination of the concepts from this cartridge, concepts from the GDPRFocusedDataDiscoveryBasic cartridge, and general full text or document metadata queries. For example, a filter looking for a date in proximity to a person in proximity to key words "birth," "death" or "hiring" may be something like this:

```
(ia:Date) w/10 (ia:Person) w/10 (birth OR death OR hiring)
```

This will search for a date close to (10 words distance) a person close to words like "birth."

## Searchable ia: terms

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

- ia:Person
- ia:Location
- ia:Organization
- ia:Address
- ia:Date
- ia:PhoneNumber

## cartridge.properties contents

```
name=GDPRFocusedDataDiscoveryAdvanced
```

```
description=Cartridge to detect the names of Persons, Locations/Addresses and Organizations/Companies for six languages
```

```
version=1.0
```

```
Technology=SystemT
```

```
languagesSupported=en,de,fr,es,zh,jp
```

```
# This is a mapping from a supportedResult name to an UIMA type
```

```
supportedResults.Person=com.ibm.systemT.Person
```

```
supportedResults.Location=com.ibm.systemT.Location
```

```
supportedResults.Organization=com.ibm.systemT.Organization
```

```
supportedResults.Address=com.ibm.systemT.Address
```

```
supportedResults.Date=com.ibm.systemT.Date
```

```
# The mapping below uses the supportedResults tag "PhoneNumber".
```

```
# The very same is used in the GDPRFocusedDataDiscoveryBasic cartridge
```

```
# This will add phone number detection for some countries that the GDPRFocusedDataDiscoveryBasic
```

```
# cartridge does not have: United States, China and Japan.
```

# For Germany, France and Spain there will be some redundancy with the results from  
# GDPRFocusedDataDiscoveryBasic. The redundancy does not have any negative effects.  
supportedResults.PhoneNumber=com.ibm.systemT.PhoneNumber.

## 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 use 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.

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.