

IDOL GDPR Package

Software Version 12.2

Technical Note



Document Release Date: February 2019
Software Release Date: February 2019

Legal notices

Copyright notice

© Copyright 2019 Micro Focus or one of its affiliates.

The only warranties for products and services of Micro Focus and its affiliates and licensors (“Micro Focus”) are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Documentation updates

The title page of this document contains the following identifying information:

- Software Version number, which indicates the software version.
- Document Release Date, which changes each time the document is updated.
- Software Release Date, which indicates the release date of this version of the software.

You can check for more recent versions of a document through the [MySupport portal](#). Many areas of the portal, including the one for documentation, require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in.

Additionally, if you subscribe to the appropriate product support service, you will receive new or updated editions of documentation. Contact your Micro Focus sales representative for details.

Support

Visit the [MySupport portal](#) to access contact information and details about the products, services, and support that Micro Focus offers.

This portal also provides customer self-solve capabilities. It gives you a fast and efficient way to access interactive technical support tools needed to manage your business. As a valued support customer, you can benefit by using the MySupport portal to:

- Search for knowledge documents of interest
- Access product documentation
- View software vulnerability alerts
- Enter into discussions with other software customers
- Download software patches
- Manage software licenses, downloads, and support contracts
- Submit and track service requests
- Contact customer support
- View information about all services that Support offers

Many areas of the portal require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in. To learn about the different access levels the portal uses, see the [Access Levels descriptions](#).

Contents

Introduction	4
Data Sources	4
Names	4
Date of Birth	5
Postal Codes	5
Addresses	5
Telephone Number	6
National Identification Number	6
Tax Identification Number (TIN)	6
Passport Number	6
Driving License	7
Medical	7
New in this Release	8
Resolved Issues	8
Country and Language Support	9
Country Codes	9
Languages	10
IDOL Eduction Grammars	12
Configure Post Processing	12
Entity Context	12
Balance Precision and Recall	13
Configure Tangible Characters	13
GDPR Grammar Reference	15
Supported National ID Numbers	19
Validated ID Numbers	20
IDOL AgentBoolean IDX	23
Send documentation feedback	25

Introduction

The IDOL GDPR Package contains tools that allow you to find personal identifiable information in your data, to help you comply with the General Data Protection Regulation (GDPR).

The IDOL GDPR Package has two types of tools:

- [IDOL Education Grammars](#) (.ecr files). IDOL Education is a tool for finding entities (small pieces of information such as names and phone numbers) in text. Education grammars contain descriptions of the entities. In some cases, this might be a list of fixed values (such as names), and in others it might be pattern matching tools that find data of a particular type (such as a set of digits that make up a phone number).

The Education GDPR grammars contain details of different kinds of personally identifiable information, to allow you to find these values in text.

- [IDOL AgentBoolean IDX](#). IDOL AgentBoolean is a method of storing entities and querying for them that uses the IDOL Agentstore component (a specially configured IDOL Content component), rather than Education. The IDX files are index files that contain the details of the entities, which you can index into the IDOL Agentstore component.

The following section describes the data sources that have been used to compile the GDPR grammars and IDX files.

Data Sources

The IDOL GDPR Package contains a variety of different kinds of entities to describe personally identifiable information that is protected by GDPR. The following sections provide some information about how this information is compiled.

For all of these types of information, as much test data is acquired as possible to test the recall metric of the algorithms. Many millions of examples are run through the grammars to ensure that all patterns in usage are covered.

Names

An international database containing over 100 million individuals across the EEA is analyzed to identify the structure and characteristics of names in each country. In doing so, extensive lists of the frequencies of occurrence of given names and family names are used to generate strong identification grammars for names.

In addition, rules are included to handle linguistic information, such as transliteration (for example, from the Cyrillic or Greek alphabets), or the use or removal of diacritic marks.

Date of Birth

A large corpus of documents from public sources is processed to analyze the occurrence and format of dates for each country subject to GDPR. In this way, coverage of all common and less-common formats is built up, while enabling a *likelihood* measure to indicate the confidence that the characters identified are a date of birth, rather than an unrelated date or other alphanumeric string.

Postal Codes

For each country of the EEA, the publications of the national Postal Services are used as the authoritative source on the postal code.

In addition, testing against widely-gathered examples allows the identification and inclusion of non-standard formats and common errors (such as mixing the letter O with the digit 0), with an appropriately adjusted likelihood measure.

Addresses

The identification of addresses consists of a number of steps, each of which is used as additional evidence that a piece of text represents a postal address. These are:

1. The format of the text.
2. The house number / street-name portion.
3. The village / town / county / region portion.
4. The postal code.

These components are not necessarily always present for a particular address, but each is taken as evidence that the text does indeed contain an address, combining to form an overall likelihood.

- Few countries have prescribed formats for addresses, while most have conventions defined by the national Postal Service that is generally adhered to, but also frequently ignored.

The IDOL Web Connector is used to gather many millions of web documents to identify candidate addresses in each applicable country. From there, the variety of formats that are used in practice across the EEA are identified. In addition, any recommendations published by the national Postal Services are also used.

- For the street-address portion, the extensive OpenStreetMap project is used, and a database of every named street in each of the countries of the EEA is obtained and analyzed. From this database, rules are derived to allow the identification of the vast majority of street-address strings.
- The de facto authority for geographical place names is the GeoNames database, with 11 million locations identified by data including country, population and type. In particular the *type* field is used to generate complete lists of populated settlements and administrative regions (such as county / department / region) for the countries that frequently use those in addresses. In addition, the names are available in different character sets and transliteration schemes to ensure internationalization.

- The patterns derived for matching Postal Codes are also used here (see [Postal Codes, on the previous page](#)).

Telephone Number

The general schemes for the creation of telephone numbers are readily available from the appropriate government department of each country. However, the formats of such numbers when written down varies considerably within a country, and even more so when numbers are referred to in a foreign document.

The strategy for creating comprehensive phone number matching grammars is centered on several key methods:

- Knowledge of the national scheme for assigning numbers.
- Databases of international and area codes in each country, obtained from authoritative sources.
- Analysis of many millions of examples of the usage of telephone numbers, obtained from a wide variety of public sources.

This final point is the most important. Only through examination of real-world usage of such numbers is the full range of formats obtained for each country.

The proximity of keywords indicating that the digits represent a telephone number is used to strengthen the likelihood of the match.

National Identification Number

Each country of the EEA has a different scheme for the use of National Identification. For countries with National ID cards, the format of the number is derived from governmental sources. In other countries, the formats of National Health, National Social Security, or National Insurance numbers are obtained from governmental sites, with the exception of a few cases in which other sources are used.

Tax Identification Number (TIN)

Each country in the European Union uses a Tax Identification Number. Grammars are used to identify these using rules laid down by the European TIN Portal, published by the European Commission.

The *strength* of the format (that is, the likelihood of false positives) and the proximity of each format to key TIN-related terms allows the calculation of a likelihood measure, where high likelihood items are stronger indicators that a TIN is present, as opposed to an unrelated number that happens to be in the same format.

Passport Number

The format of the national passport numbers is not as widely available as other such numbers. However, authoritative government documents are acquired for the formats of passport numbers in 16 of the countries of the EEA.

In other cases, non-governmental sources and the examination of examples have allowed grammars to be created for each country. In all cases, the presence of keywords and phrases in appropriate languages in proximity to the number are used to increase the likelihood of the match and to reduce the number of false positives.

In addition, grammars to identify Machine-Readable travel documents such as the MROTD and MRP have been added.

Driving License

As with passport numbers, not all governments have published the scheme used in the numbering of Driving Licenses. The format of the number is obtained for the majority of relevant countries, with the remainder derived from secondary sources and from analysis of example numbers.

Medical

Documents that contain mention of medical procedures or conditions are identified with the Medical categories, available in each of the languages subject to GDPR. The categories are generated from the Medical Subject Headings (MeSH) taxonomy published by the United States National Library of Medicine using the C hierarchy (diseases and conditions).

New in this Release

This section describes the enhancements to the IDOL GDPR Package in version 12.2.

- Significantly improved processing speed for Education when extracting names (with the `name.ecr` grammar file), addresses (`address.ecr`), and telephone numbers (`telephone.ecr`). As a result of these changes, it is now essential to configure a Lua post processing task to run the script `gdpr_postprocessing.lua`.
- The IDOL GDPR Package includes new grammar files containing entities that match a specific type of data for all supported countries and languages. For example, `combined_address.ecr` contains an entity named `gdpr/address/all` that matches a postal address for any supported country. These combined entities provide a significant improvement in processing speed when you extract matches for all countries or languages. For more information about these grammar files, see the [Grammar Reference](#).
- The Education grammar for telephone numbers has been improved to reduce the number of spurious matches. The grammar now considers any non-whitespace or non-numeric character to be a word separator. Micro Focus recommends that you include the hyphen character (-) in the list of tangible characters when using this grammar.
- The Education grammar for postcodes has been improved so that British post codes are extracted even if the space is omitted.
- The Education grammar for passport numbers has been updated to match passport numbers in additional contexts.
- The example Education configuration has been modified so that matches located without context are returned if they are successfully validated (some ID numbers can be validated by calculating a checksum). A valid checksum indicates that a match is more likely (but not guaranteed) to be genuine.

Resolved Issues

This section lists the resolved issues in the IDOL GDPR Package version 12.2.

- The Education grammars now match newlines regardless of whether the input text has Windows (CR LF) or Unix (LF) line endings.

Country and Language Support

The IDOL GDPR Package contains grammars and IDX files that apply to data from several countries and languages.

Country Codes

For data that corresponds to a particular country, the Education grammars identify each country by using the ISO 3166-1 alpha-2 country codes. The following countries are supported:

Country Code	Country
at	Austria
be	Belgium
bg	Bulgaria
cy	Cyprus
cz	Czech Republic
de	Germany
dk	Denmark
ee	Estonia
es	Spain
fi	Finland
fr	France
gb	United Kingdom (England, Wales, Scotland, and Northern Ireland)
gr	Greece
hr	Croatia
hu	Hungary
ie	Ireland
is	Iceland
it	Italy

Country Code	Country
li	Liechtenstein
lt	Lithuania
lu	Luxembourg
lv	Latvia
mt	Malta
nl	Netherlands
no	Norway
pl	Poland
pt	Portugal
ro	Romania
se	Sweden
si	Slovenia
sk	Slovakia

Languages

For data that corresponds to a particular language, the Eduction grammars and AgentBoolean IDX files identify each language by using the ISO 639-2/B language codes. The following languages are supported:

Language Code	Language
bul	Bulgarian
cat	Catalan
cze	Czech
dan	Danish
dut	Dutch
eng	English
est	Estonian
fin	Finnish

Language Code	Language
fre	French
ger	German
gle	Irish
gre	Greek
hrv	Croatian
hun	Hungarian
ice	Icelandic
ita	Italian
lav	Latvian
lit	Lithuanian
mlt	Maltese
nor	Norwegian
pol	Polish
por	Portuguese
rum	Romanian
slo	Slovak
slv	Slovenian
spa	Spanish
swe	Swedish

IDOL Education Grammars

The following section describes the GDPR Education grammars available in the IDOL GDPR Package.

You can use these grammars with IDOL Education, by using Education Server, the `edktool` command-line utility, or the Education SDK. For more information, refer to the *IDOL Education User Guide* and the *Education SDK Programming Guide*.

IMPORTANT:

To use the GDPR grammars with Education, you must have a license that enables them. To obtain a license, contact Micro Focus Support.

Configure Post Processing

When you use the GDPR Education grammars it is essential to configure a Lua post-processing task to run the script `gdpr_postprocessing.lua`. If you do not run this script, you might encounter unexpected behavior.

Add a post-processing task to your Education configuration. For example:

```
[PostProcessingTasks]
NumTasks=1
Task0=MyPostProcessingSection

[MyPostProcessingSection]
Type=Lua
Script=scripts/gdpr_postprocessing.lua
Entities=gdpr/*
```

For more information about configuring post-processing tasks, refer to the *Education User Guide*.

Entity Context

Some of the entities are available in two versions, with and without context. The context-based entities match the entity when it occurs in an easily identifiable location in text. For example, it might match a telephone number that occurs next to the prefix **Phone**:

The entities that do not have context attempt to match the entity wherever it occurs. This version might over-match significantly (that is, it is likely to return values that are similar to the entity patterns, such as a number that is not a telephone number). However, it also reduces the number of false negatives (that is, it misses fewer matches).

You can configure Education to use both versions of an entity; matches located with context are given a higher score in the results.

Balance Precision and Recall

In many cases, Education is able to locate entities that are ambiguous, such as a postal code which is simply a five-digit number. In some situations it is desirable to match as many entities as possible ("high recall") and in others only entities with a high likelihood of being a useful match ("high precision"). So that you can filter the results produced by Education, each match is given a 'score' value.

As described in [Entity Context](#), matches located by an entity that requires context are assigned higher scores than matches located by the corresponding entity without context. Most matches extracted without context have a score of 0.4. For example, a context-free date ("January 18, 1998") might be returned by the Date Of Birth entity with a score of 0.4. But with context to suggest that it is indeed a date of birth ("DOB: January 18, 1998"), the score should be above 0.5.

The GDPR post-processing script (see [Configure Post Processing, on the previous page](#)) includes a step to validate matches (some ID numbers can be validated by calculating a checksum). The script increases the score of matches that have valid checksums, because this is an indication that the match is more likely to be genuine. Any match that has an invalid checksum is immediately discarded because it cannot be genuine.

When you configure Education, use the parameters `MinScore` and `PostProcessThreshold` to achieve the desired balance between precision and recall. Education discards any match with a score lower than `MinScore`. Matches with scores that meet or exceed `MinScore` are then processed by post-processing tasks. After post-processing has finished, Education discards any match with a score lower than `PostProcessThreshold`.

In the example configuration that is included with the IDOL GDPR Package, `MinScore` is set to 0.4 and `PostProcessThreshold` is set to 0.5. These values have been chosen to return results only if they have a relatively high likelihood of being a useful match. Any match that is located without context can proceed to post-processing, but, unless its score is increased through successful validation, it is then discarded. If you prefer to maximize recall rather than precision, you can reduce or remove these thresholds.

For more information about Education configuration parameters, refer to the *Education User Guide*.

Configure Tangible Characters

`TangibleCharacters` is a configuration parameter that you can set when using the Education SDK, the Education Server, or the Education command-line utility (`edktool`). It specifies a list of characters to treat as part of a word, rather than as word boundaries.

Some of the entities in the GDPR Package Education Grammars require tangible characters to be set in order to perform correctly (see the descriptions of the entities in [GDPR Grammar Reference, on page 15](#)).

When you use Education to search for matches, `TangibleCharacters` applies across all of your chosen entities. If you use multiple entities that have different recommended tangible character sets, you might need to take some extra steps. For example:

- If you are using the Education SDK, create a separate EDK engine for each distinct set of tangible characters, and configure the tangible characters for the engine using the appropriate API call:

C `EdkSetTangibleCharacters`

Java `EDKEngine.setTangibleCharacters`

After configuring an engine with the correct tangible characters, you can add the relevant entities. You will need to create a session from each engine to process your input text.

- If you are using an Education Server, send a separate action (`EduceFromText` or `EduceFromFile`) for each distinct set of tangible characters. In each action, set the `TangibleCharacters` and `Entities` action parameters to specify which set of tangible characters and which entities to use.
- If you are using the command line `edktool`, create a separate configuration file for each distinct set of tangible characters and associated entities, and process your input text once with each configuration file.

For more information about the `TangibleCharacters` configuration parameter, refer to the *Education User Guide*.

GDPR Grammar Reference

The following table describes the grammar files that are available in the GDPR package, and the entities that each provides.

In the entity names:

- the abbreviation CC refers to a two-letter country code. For a list of available country codes, see [Country Codes, on page 9](#).
- the abbreviation LLL refers to a three-letter language code. For a list of available languages, see [Languages, on page 10](#).

TIP:

You can use the Education parameter `EntityN` to specify which entities you want to extract. This parameter accepts wildcards, so you can extract entities of a specific type for all supported countries or languages. For example, to match postal addresses for all countries specify a value of `gdpr/address/??`. To match dates of birth in all languages, specify `gdpr/date/dob/???`.

File	Entity	Description
address.ecr	gdpr/address/CC	A postal address.
date.ecr	gdpr/date/dob/LLL	A date of birth, written numerically or using words.
	gdpr/date/nocontext/LLL	A calendar date, written numerically or using words, without context.
driving.ecr	gdpr/driving/CC	A driving license number with context.
	gdpr/driving/nocontext/CC	A driving license number, without context.

File	Entity	Description
health.ecr	gdpr/health/ehic/CC	An EHIC personal identification number with context.
	gdpr/health/ehic/nocontext/CC	An EHIC personal identification number without context.
	gdpr/health/tarjeta_sanitaria/es	A Spanish health insurance card number with context.
	gdpr/health/nhs/gb	A British NHS number with context.
	gdpr/health/carte_vitale/fr	A French Carte Vitale number with context.
mrttd.ecr	gdpr/mrttd/mrp	A machine readable passport.
	gdpr/mrttd/mrotd/td1	A machine readable TD1-size travel document.
name.ecr	gdpr/name/CC	A full personal name.
national_id.ecr	gdpr/id/CC	A national identity number with context. For information about the supported ID numbers, see Supported National ID Numbers, on page 19 .
	gdpr/id/nocontext/CC	A national identity number without context. For information about the supported ID numbers, see Supported National ID Numbers, on page 19 .
passport.ecr	gdpr/passport/CC	A passport number with context.
	gdpr/passport/nocontext/CC	A passport number without context.
postcode.ecr	gdpr/postcode/CC	A postal code with context.
	gdpr/postcode/nocontext/CC	A postal code without context.
telephone.ecr	gdpr/telephone/CC	A telephone number with context.
		<p>NOTE: To ensure that this entity performs correctly, set your</p>

File	Entity	Description
		<p>TangibleCharacters configuration to include the following characters: ()+-. For more information, see Configure Tangible Characters, on page 13.</p>
	gdpr/telephone/nocontext/CC	<p>A telephone number without context.</p> <p>NOTE: To ensure that this entity performs correctly, set your TangibleCharacters configuration to include the following characters: ()+-. For more information, see Configure Tangible Characters, on page 13.</p>
tin.ecr	gdpr/tin/CC	A tax identification number with context.
	gdpr/tin/nocontext/CC	A tax identification number without context.

File	Entity	Description
combined_address.ecr	gdpr/address/all	<p>These entities match addresses, dates, driving license numbers, and so on, from any supported country or language.</p> <p>For example:</p> <ul style="list-style-type: none"> Using <code>gdpr/address/all</code> from <code>combined_address.ecr</code> matches a postal address from any country. This is similar to using the <code>address.ecr</code> grammar file and extracting <code>gdpr/address/??</code>. Using <code>gdpr/date/dob/all</code> from <code>combined_date.ecr</code> matches a date of birth written numerically or using words in any language. This is similar to using the <code>date.ecr</code> grammar file and extracting <code>gdpr/date/dob/???</code>. <p>The combined (<code>/all</code>) entities provide a significant improvement in processing speed when you extract matches for all countries or languages.</p> <p>You must run the script <code>gdpr_postprocessing.lua</code> as a post-processing task (see Configure Post Processing, on page 12). Running the script ensures that the entity names returned by Education contain the relevant country code or language code. For example, if a UK postal address is found, the entity name in the returned match is still <code>gdpr/address/gb</code>, and not <code>gdpr/address/all</code>.</p> <p>The combined grammar files might produce fewer matches, because only a single match is returned in cases where the same characters in the input text would match multiple countries or languages.</p>
combined_date.ecr	gdpr/date/dob/all	
	gdpr/date/nocontext/all	
combined_driving.ecr	gdpr/driving/all	
	gdpr/driving/nocontext/all	
combined_health.ecr	gdpr/health/ehic/all	
	gdpr/health/ehic/nocontext/all	
combined_name.ecr	gdpr/name/all	
combined_national_id.ecr	gdpr/id/all	
	gdpr/id/nocontext/all	
combined_passport.ecr	gdpr/passport/all	
	gdpr/passport/nocontext/all	
combined_postcode.ecr	gdpr/postcode/all	
	gdpr/postcode/nocontext/all	
combined_telephone.ecr	gdpr/telephone/all	
	gdpr/telephone/nocontext/all	
combined_tin.ecr	gdpr/tin/all	
	gdpr/tin/nocontext/all	

Supported National ID Numbers

The following table lists the national ID numbers that are supported by the `gdpr/id/CC` and `gdpr/id/nocontext/CC` Education entities.

Country	Supported national identity numbers	Example context
Austria	SSN (social security number) CRR	ASVG
Belgium	NRN (numéro de registre national)	numéro national
Bulgaria	EGN (Uniform Civil Number)	EGN
Croatia	OIB (Osobni identifikacijski broj)	OIB
Cyprus	Identity card number	Αριθμός ταυτότητας
Czech republic	Rodné číslo	rodné číslo
Denmark	CPR	legitimation
Estonia	IK (isikukood)	Isikukood
Finland	Henkilötunnus (Personal identity code)	Henkilötunnus
France	INSEE code	Code INSEE
Germany	National ID serial number	Personalausweis
Greece	National ID card AMKA (social security number)	AMKA
Hungary	Personal Identification Number ID card number	Nemzeti személyazonosító jel
Iceland	Kennitala	Kennitala
Ireland	PPSN (personal public service number)	PPSN
Italy	Codice Fiscale	codice fiscale
Latvia	Personas kods	personas kods
Liechtenstein	Identitätskarte	Personalausweis
Lithuania	Asmens Kodas	asmens kodas
Luxembourg	National ID card number Identity card number	Steuernummer

Malta	ID card number	ID card
Netherlands	BSN	legitimatiebewijs
Norway	Fødselsnummer D-nummer H-nummer FH-nummer	Fødselsnummer
Poland	PESEL	PESEL
Portugal	Número de identificação civil Cartão de cidadão number Número de Identificação de Segurança Social	NIC
Romania	Cod Numeric Personal	CNP
Slovakia	Rodné číslo ID card number	rodné číslo
Slovenia	Enotna matična številka občana	EMŠO
Spain	DNI NIE	DNI
Sweden	Personnummer Samordningsnummer	personnummer
UK	National Insurance Number	National Insurance Number

Validated ID Numbers

The script `gdpr_postprocessing.lua` (see [Configure Post Processing, on page 12](#)) includes steps to validate ID numbers that are found by Education. This improves accuracy by discarding results that match the pattern for a valid ID number, but cannot be genuine because they have an invalid checksum. The script increases the score for matches that have a valid checksum, because this is an indication that the match is more likely to be genuine.

The following tables list the entities that are validated.

Health ID numbers (<code>health.ecr</code>)	
<code>gdpr/health/nhs/gb</code>	
<code>gdpr/health/carte_vitale/fr</code>	Validated using the INSEE checksum

National ID numbers (national_id.ecr)		
gdpr/id/at	gdpr/id/nocontext/at	Only the SSN component is validated.
gdpr/id/be	gdpr/id/nocontext/be	
gdpr/id/bg	gdpr/id/nocontext/bg	
gdpr/id/cz	gdpr/id/nocontext/cz	
gdpr/id/ee	gdpr/id/nocontext/ee	
gdpr/id/es	gdpr/id/nocontext/es	
gdpr/id/fi	gdpr/id/nocontext/fi	
gdpr/id/fr	gdpr/id/nocontext/fr	
gdpr/id/gr	gdpr/id/nocontext/gr	Only the AMKA component is validated.
gdpr/id/hr	gdpr/id/nocontext/hr	
gdpr/id/hu	gdpr/id/nocontext/hu	Only the PIN component is validated.
gdpr/id/ie	gdpr/id/nocontext/ie	
gdpr/id/is	gdpr/id/nocontext/is	
gdpr/id/it	gdpr/id/nocontext/it	
gdpr/id/lt	gdpr/id/nocontext/lt	
gdpr/id/lu	gdpr/id/nocontext/lu	
gdpr/id/nl	gdpr/id/nocontext/nl	
gdpr/id/no	gdpr/id/nocontext/no	
gdpr/id/pl	gdpr/id/nocontext/pl	
gdpr/id/pt	gdpr/id/nocontext/pt	
gdpr/id/ro	gdpr/id/nocontext/ro	
gdpr/id/si	gdpr/id/nocontext/si	
gdpr/id/se	gdpr/id/nocontext/se	
gdpr/id/sk	gdpr/id/nocontext/sk	Only the Rodné číslo component is validated.

Tax ID numbers (tin.ecr)	
gdpr/tin/at	gdpr/tin/nocontext/at

gdpr/tin/be	gdpr/tin/nocontext/be
gdpr/tin/bg	gdpr/tin/nocontext/bg
gdpr/tin/cy	gdpr/tin/nocontext/cy
gdpr/tin/de	gdpr/tin/nocontext/de
gdpr/tin/dk	gdpr/tin/nocontext/dk
gdpr/tin/ee	gdpr/tin/nocontext/ee
gdpr/tin/es	gdpr/tin/nocontext/es
gdpr/tin/fi	gdpr/tin/nocontext/fi
gdpr/tin/fr	gdpr/tin/nocontext/fr
gdpr/tin/hr	gdpr/tin/nocontext/hr
gdpr/tin/hu	gdpr/tin/nocontext/hu
gdpr/tin/ie	gdpr/tin/nocontext/ie
gdpr/tin/it	gdpr/tin/nocontext/it
gdpr/tin/lt	gdpr/tin/nocontext/lt
gdpr/tin/lu	gdpr/tin/nocontext/lu
gdpr/tin/mt	gdpr/tin/nocontext/mt
gdpr/tin/nl	gdpr/tin/nocontext/nl
gdpr/tin/pl	gdpr/tin/nocontext/pl
gdpr/tin/pt	gdpr/tin/nocontext/pt
gdpr/tin/se	gdpr/tin/nocontext/se
gdpr/tin/si	gdpr/tin/nocontext/si
gdpr/tin/sk	gdpr/tin/nocontext/sk

IDOL AgentBoolean IDX

IDOL AgentBoolean provides another way of finding pieces of information in text. In this case, you index the entities that you want to find into an IDOL Agentstore component.

The IDOL Agentstore component is a specially configured IDOL Content component. It uses IDOL AgentBoolean queries for entity matching.

When you use AgentBoolean for entity matching, each entity becomes a document in Agentstore. You then send a piece of text as a query to Agentstore, and it returns the entity documents that match the text.

The IDOL GDPR Package contains several IDX documents that describe entities for medical data, which you can use as another tool to find data that is protected by GDPR. The package also contains example Agentstore configuration files to allow you to set up your Agentstore component more easily.

There is an IDX file for each of the supported languages (see [Languages, on page 10](#)).

After you configure and set up your Agentstore, you can index the IDX documents and use Agentstore for entity matching.

For more information about how to set up and use IDOL querying, refer to the *IDOL Server Administration Guide* and the *IDOL Content Component Reference*.

Send documentation feedback

If you have comments about this document, you can [contact the documentation team](#) by email. If an email client is configured on this system, click the link above and an email window opens with the following information in the subject line:

Feedback on Technical Note (Micro Focus IDOL GDPR Package 12.2)

Add your feedback to the email and click **Send**.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to swpdl.idoldocsfeedback@microfocus.com.

We appreciate your feedback!