IDOL PII Package

Software Version 12.7

Technical Note

Document Release Date: October 2020
Software Release Date: October 2020

Legal notices

Copyright notice

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.

To check for updated documentation, visit https://www.microfocus.com/support-and-services/documentation/.

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. If you need an account, 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

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

The IDOL PII Package uses IDOL Eduction Grammars (.ecr files).

IDOL Eduction is a tool for finding entities (small pieces of information such as names and phone numbers) in text. Eduction 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 Eduction grammars included in the IDOL PII Package describe different kinds of personally identifiable information, so that you can find these in your data.

Data Sources

The IDOL PII Package contains a variety of different kinds of entities to describe personally identifiable information that is protected by regulations such as 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 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 supported country. 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, 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 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 supported countries 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, above).

Telephone Number

The general schemes for the creation of telephone numbers and fax 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 or fax number is used to strengthen the likelihood of the match.

National Identification Number

Each country 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 the majority of supported countries.

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 supported languages. 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 PII Package in version 12.7.

- The IDOL PII Package now includes resources for South Africa. A complete set of entities are available to extract information including addresses and postcodes, dates, driving license numbers, names, nationality, national ID numbers, passport numbers, telephone numbers, and tax identification numbers.
- The IDOL PII Package now includes resources for Taiwan. A complete set of entities are available to extract information including addresses and postcodes, dates, driving license numbers, names, nationality, national ID numbers, health numbers, passport numbers, telephone numbers, and tax identification numbers.
- A new grammar, `device_id` has been added to match various device identifiers. This grammar is available in ECR and EJR formats. For more details, see Eduction Grammar Reference, on page 20.
- The address grammar now returns the `PO_BOX` component for all countries (previously, only UK and USA had this component). In addition, the following countries now detect the post office name, with the `POST_OFFICE` component: Canada, Italy, Lithuania, Norway, New Zealand, Portugal, Taiwan, and Japan.
- The address grammar has been improved to reduce false positive matches. In particular, in the recommended configuration, the grammar no longer matches an unknown street with an unknown city. Either the street name or the city must belong to one of the known lists.
- The standard PII grammars now detect additional types of spaces in input text in all the places where previously regular spaces were expected. This change adds detection for U+00A0 (no break space), U+2007 (figure space), and U+3000 (ideographic space). Where these spaces are detected in input text, they are normalized to regular spaces.

  Similarly, the PII grammars now detect and normalize additional apostrophe characters in places where a regular apostrophe was expected. This change adds detection for U+2019 (right single quote), and U+FF07 (full-width apostrophe).
- The national ID grammar now matches national IDs for Bahrain, Dominican Republic, Egypt, Indonesia, Mexico, Pakistan, and Russia.
- The `name_cjkvt` grammar now has additional entities to match Latin-only and CJKVT-only versions of full names.
- The name and address grammars now have additional landmark entities for the full name or address.
- For the TIN and national ID entities, you can now enable ambiguous entity matching by setting `ambiguous_tin_id_entities=true` in the `pii_postprocessing.lua` script. This option returns multiple possible country matches. By default, post-processing returns only one country, which is more efficient.
- Synonyms have been added to street components of the address entities of many countries. In

these cases, when a synonym is detected, the main headword is returned in the normalized text. For example, Oxford St normalizes to Oxford Street in address entity matches for English-speaking countries.

- The address grammar now returns additional components for county (for the UK) and province (for Canada).

Resolved Issues

This section lists the resolved issues in the IDOL PII Package version 12.7.

- Offsets and offset lengths returned for components were different for EJR grammars compared to their equivalent ECR grammar.
- EJR grammars could return more components than equivalent ECR grammar.
- When using Eduction in CFS or NiFi, Lua post-processing scripts that renamed the matching entity could drop matches, or return the match under an unexpected field name (such as "_"). For example, the post-processing scripts shipped in the IDOL PII Package rename entities from the combined grammars ending "/all" to correspond to the matched language.
- Post-processing replaced the normalized text with empty space for pii/name/given_ name/nocontext/all or pii/name/surname/nocontext/all entity matches.
- When matching nocontext name components other than given name or surname (such as pii/name/pre_title/all), post-processing could log an error "Error during call to lua function 'processmatch' in script '.../pii_postprocessing.lua': Parameter 2 had the wrong type".
- In post-processing for TIN and national ID, an issue with the lua script resulted in some false positive matches when scanning for alternative matches after an initial context or nocontext match had failed.
- For TIN and national ID, nocontext matches for countries that did not have a checksum algorithm could receive a score boost in post-processing. Now, these matches receive a score boost only if they are confirmed by checksum.
- In the `pii_postprocessing.lua` scripts, `print()` was used instead of `error()` or `assert()` in some cases, which could result in errors not being correctly returned when using the Eduction SDK.
- Scoring in the `name_cjkvt` grammar did not accurately reflect common Japanese given names.

Country and Language Support

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

Country Codes

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

Country Code	Country
at	Austria
au	Australia
be	Belgium
bg	Bulgaria
br	Brazil
ca	Canada
ch	Switzerland
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

Country Code	Country
hu	Hungary
ie	Ireland
is	Iceland
it	Italy
jp[1]	Japan
li	Liechtenstein
lt	Lithuania
lu	Luxembourg
lv	Latvia
mt	Malta
nl	Netherlands
no	Norway
nz	New Zealand
pl	Poland
pt	Portugal
ro	Romania
se	Sweden
si	Slovenia
sk	Slovakia
tr	Turkey
tw[2]	Taiwan
us	United States of America
za	South Africa

For national IDs, the following additional country codes are supported (see Supported National ID Numbers, on page 47)

[1]This country is available only in CJKVT grammars.
[2]This country is available only in CJKVT grammars.

Country Code	Country
ae	United Arab Emirates
ar	Argentina
bh	Bahrain
co	Colombia
cn	China
do	Dominican Republic
eg	Egypt
hk	Hong Kong
id	Indonesia
in	India
mx	Mexico
my	Malaysia
pk	Pakistan
ru	Russia
sg	Singapore
th	Thailand

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
afr	Afrikaans
bul	Bulgarian
cat	Catalan
chi[1]	Chinese

[1]This language is available only in CJKVT grammars.

Language Code	Language
cze	Czech
dan	Danish
dut	Dutch
eng	English
est	Estonian
fin	Finnish
fre	French
ger	German
gle	Irish
gre	Greek
hrv	Croatian
hun	Hungarian
ice	Icelandic
ita	Italian
jpn[1]	Japanese
lav	Latvian
lit	Lithuanian
mlt	Maltese
nor	Norwegian
nso	Northern Sotho
pol	Polish
por	Portuguese
roh	Romansh
rum	Romanian
slo	Slovak
slv	Slovenian

[1]This language is available only in CJKVT grammars.

Language Code	Language
spa	Spanish
ssw	Swati
swe	Swedish
tso	Tsonga
tur	Turkish
ven	Venda
xho	Xhosa
zul	Zulu

IDOL Eduction Grammars

The following section describes the Eduction grammars available in the IDOL PII Package.

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

> **IMPORTANT:** To use the Eduction grammars in the IDOL PII Package, you must have a license that enables them. To obtain a license, contact Micro Focus Support.

The IDOL PII Package includes a default configuration file, which includes the basic required settings that you need to use the PII grammars.

> **NOTE:** If you create your own configuration file, you must include some of the settings in the default configuration file, such as post-processing and Eduction *components* (see Configure Post Processing, below).

Configure Post Processing

When you use the IDOL PII Package Eduction grammars it is essential to configure a Lua post-processing task to run the script `pii_postprocessing.lua`. This script contains post-processing to improve results for various entities, such as stop list filtering, entity name mapping for combined grammars (see Combined Entities, on page 35), ambiguous landmark detection (see Ambiguous Entities, on page 83) and checksum validation (see Validated ID Numbers, on page 79).

> **IMPORTANT:** If you do not run this script, you might encounter unexpected behavior.

The default configuration file provided in the IDOL PII Package includes a suitable post-processing task. If you use a different configuration, you must add the post-processing task to your Eduction configuration. For example:

```
[Eduction]
PostProcessingTask0=MyPostProcessingSection

[MyPostProcessingSection]
Type=Lua
Script=scripts/pii_postprocessing.lua
Entities=pii/*,gdpr/*
```

> **IMPORTANT:** The post-processing script requires Eduction components (see Components, on page 39). The default PII configuration file enables components. If you use a custom configuration file you must set the `EnableComponents` parameter to `True` to return components.

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

Configure Pre-Filtering

Pre-filtering allows the IDOL PII Package to run a quick initial check to find potential matches in your input text. It then selects match windows around these potential matches, reducing the amount of text that it must match against your grammars. This process can improve the performance in certain cases.

Micro Focus recommends that you use the following pre-filtering configuration with the `address.ecr` and `combined_address.ecr` grammars.

```
[Eduction]
PrefilterTask0=AddressPrefilter

[AddressPrefilter]
Regex=\d{1,7}
WindowCharsBeforeMatch=100
WindowCharsAfterMatch=100
```

> **NOTE:** Pre-filter tasks run for all configured entities, so you must configure it only for the appropriate entities to ensure that it does not affect the results for other entities.

For more information about pre-filtering, refer to the *Eduction User and Programming 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 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 Eduction to use both versions of an entity; matches located with context are given a higher score in the results.

When you have data in tables, the context for an entity might not occur next to the entity value. For example, you might have a table with columns titled **name** and **date of birth**, but the values themselves do not occur next to these headers.

In this case, you can use Eduction table extraction to extract entities according to the landmarks detected in the table headers. For example, you can configure Eduction so that if it finds a table heading that matches the landmark **date of birth**, it extracts dates from that column.

For more information about how to configure table extraction, refer to the *Eduction User and Programming Guide*.

ECR and EJR Grammars

Some grammars are available in two formats, ECR and EJR. In these cases, both formats contain the same entities for extraction, and the format that you use depends on your input data.

EJR files are performance-optimized for cases where the expected match density in your input text is low. Micro Focus recommends that you use EJR files when you expect less than 10% of the input text to be valid matches. In all other cases, use the ECR files.

When you use EJR grammars, you must run them in a separate matching engine to any ECR grammars, although you can run multiple EJR grammars in the same engine.

For example, the following configuration is allowed:

```
ResourceFiles=passport.ejr,date.ejr
```

You cannot set `ResourceFiles=passport.ejr,date.ecr`.

Balance Precision and Recall

In many cases, Eduction 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"). Each match is given a score value so that you can filter the results.

As described in Entity Context, on the previous page, 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 a 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 PII post-processing script (see Configure Post Processing, on page 16) includes a step to validate matches (for example, it can validate some ID numbers 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 Eduction, use the parameters `MinScore` and `PostProcessThreshold` to achieve the desired balance between precision and recall. Eduction 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, Eduction discards any match with a score lower than `PostProcessThreshold`.

In the example configuration that is included with the IDOL PII 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 Eduction configuration parameters, refer to the *Eduction User and Programming Guide*.

Configure Tangible Characters

`TangibleCharacters` is a configuration parameter that you can set when using the Eduction SDK, the Eduction Server, or the Eduction 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 IDOL PII Package Eduction Grammars require tangible characters to be set in order to perform correctly (see the descriptions of the entities in Eduction Grammar Reference, on the next page).

When you use Eduction 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 Eduction 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 Eduction 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 *Eduction User Guide*.

Customize Stop Lists

The IDOL PII Package post-processing script (see Configure Post Processing, on page 16) uses stop lists to discard matches that are likely to be false positives. You can add entries to the stop lists, or remove entries, by modifying the following files.

- `scripts/address_stoplist.lua` contains a list of common words that are likely to indicate a false positive when returned as the `STREET` or `CITY` component of an address match.
- `scripts/names_stoplist.lua` contains two stop lists to discard names. In the first stop list, each component is plausible but the entire match is likely to be a false positive, for example "Christian Church" or "Norman Conquest". The second stop list contains common words that are

likely to indicate a false positive when returned as either the FORENAME or SURNAME component of a name match. The stop lists in this file can be customized such that a name can be considered a false positive in one country but not another.

Eduction Grammar Reference

The following tables describe the grammar files that are available in the IDOL PII Package, and the entities that each provides.

Some grammars are available in two formats, ECR and EJR. For more information about which to use, see ECR and EJR Grammars, on page 18.

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 11.
- the abbreviation LLL refers to a three-letter language code. For a list of available languages, see Languages, on page 13.

TIP: You can use the Eduction 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 `pii/address/??`. To match dates of birth in all languages, specify `pii/date/dob/context/???`.

Some grammars have a CJVTK (Chinese, Japanese, Korean, Vietnamese, and Thai) version, for matching entities in Japanese. This grammar is separate for performance reasons. CJKVT languages do not have spaces in the text to separate words, so additional processing is required for sentence breaking.

The CJKVT grammars generally match Kanji and Romanized versions of the text where appropriate, as well as half-width and full-width characters (in the output, full-width forms are normalized to half-width).

NOTE: The IDOL PII Package is backwards-compatible with the IDOL GDPR package. You can continue to use existing configurations that use entity names such as `gdpr/address/CC` or `gdpr/telephone/CC`. These entities are similar to the corresponding `pii/*` entity, but are limited to countries in the GDPR region. However, Micro Focus recommends that you use the `pii/*` entities instead, so that Eduction extracts matches for all supported countries.

address.ecr

Entity	Description
pii/address/CC	A postal address. In general, a score of one is given to an address that includes a numbered, common format street address (for

Entity	Description
	example "23 North Road"), a known city (for example "London"), and a postal code in a viable format for the country (for example "SW1A 2AA"). Deviations from this form lead to score penalties. The ordering of these elements varies by country. Micro Focus recommends that you use pre-filtering to improve the performance for this grammar. See Configure Pre-Filtering, on page 17. Example matches: "Schlosshoferstrasse 20, 1210 Vienna", "Avenida Juan Xxiii 20, 41006, Sevilla", "Abidei Hurriyet Cd Taner Palas Han 9 Kat:7 Dayre 9, 34437 Istanbul", "162-168 Regent Street, London, W1B 5TG". This entity returns the addresses in a normalized format by default. The normalized form standardizes apartment and house numbers, removes additional punctuation, and converts the text to uppercase. For example ABIDEI HURRIYET CD TANER PALAS APT 9, KAT:7, D:9, 34437 ISTANBUL. The exact order depends on the country. You can turn off normalization by setting `normalize_addresses=false` in the `address_stoplist.lua` script. This option can improve performance when you do not need normalization. This entity returns components. See Components, on page 39.
pii/address/landmark/CC	A postal address landmark. For example "Address".
pii/address/streetlocation/context/CC	A street location (house number and street name), with context. For example "Address: 123, Mill Road".
pii/address/streetlocation/nocontext/CC	A street location (house number and street name), without context. For example "123, Mill Road".
pii/address/streetlocation/landmark/CC	A street location landmark. For example "Address"
pii/address/city/context/CC	A city or town, with context. For example "City: London".
pii/address/city/nocontext/CC	A city or town, without context. For example "London".
pii/address/city/landmark/CC	A city or town landmark. For example "City".
pii/address/postcode/context/CC	A postal code, with context. For example "Postcode: CB4 0WZ".
pii/address/postcode/nocontext/CC	A postal code, without context. For example "CB4 0WZ".

Entity	Description
pii/address/postcode/landmark/CC	A postal code landmark. For example "Postcode".
pii/address/country/context/CC	A country, with context. For example "Country: United Kingdom".
pii/address/country/nocontext/CC	A country, without context. For example "United Kingdom".
pii/address/country/landmark/CC	A country landmark. For example "Country".

address_cjkvt.ecr

Entity	Description
pii/address/CC	A postal address. In general, for Japan, a score of one is given to an address that includes a numbered, common format street address (for example "四丁目 1番 2-34号"), a city or ward (for example 津島市), and a postal code in a viable format for the country (for example "１２３-４５６７"). For Taiwan, a score of one is given to an address that includes a numbered street address, a known township, district, city, or county, and a valid postal code. Deviations from these forms lead to score penalties. Micro Focus recommends that you use pre-filtering to improve the performance for this grammar. See Configure Pre-Filtering, on page 17 Example matches: "日本、〒１２３-４５６７神奈川県津島市城南区月形町八重洲四丁目 1番 2-34号", "1-2-34, Yaesu 4-Chome, Nanae, Atsuta, Hekinan, Kagoshima, 123-4567, Japan", "１０６０３台北市大安區金山南路 2段 55號", "No.55, Sec. 2, Jinshan S. Rd., Daan Dist., Taipei City 10603". This entity returns the addresses in a normalized format. The normalized form standardizes apartment and house numbers, removes additional punctuation, and for Romanized text, it converts the text to uppercase. CJKVT native script is not normalized to ASCII, and Romanized text is not normalized to CJKVT native script.

Entity	Description
	You can turn off normalization by setting `normalize_addresses=false` in the `address_stoplist.lua` script. This option can improve performance when you do not need normalization. This entity returns components. See Components, on page 39.
pii/address/landmark/CC	A postal address landmark. For example "住所".
pii/address/streetlocation/contextCC	An address first line, with context. For example "住所: 八重洲 四丁目 1番 2-34号" or "Address: No.55, Sec. 2, Jinshan S. Rd.".
pii/address/streetlocation/nocontext/cjkvt/CC	An address first line in CJKVT native script, without context. For example "八重洲四丁目 1番 2-34号", or "金山南路 2段 55號".
pii/address/streetlocation/nocontext/latin/CC	An address first line in romanized text, without context. For example "1-2-34, Yaesu 4-Chome", or "No.55, Sec. 2, Jinshan S. Rd.".
pii/address/streetlocation/nocontext/CC	An address first line in CJKVT native script or romanized text, without context. For example 八重洲四丁目 1番 2-34号" or "1-2-34, Yaesu 4-Chome".
pii/address/streetlocation/landmark/CC	An address first line landmark. For example "住所", "住址", or "Address".
pii/address/settlement/context/CC	A settlement, with context. For example, in Japan a town or city, or in Taiwan a district (區 qū) or township (鎮 zhèn/鄉 xiāng). For example "市区町村: 津島市城南区月形町", "鄉鎮市區:板橋區", or "City/Ward/Town/Village: Nanae, Atsuta, Hekinan".
pii/address/settlement/nocontext/cjkvt/CC	A settlement in CJKVT native script, without context. For example "津島市城南区月形町", or "板橋區".
pii/address/settlement/nocontext/latin/CC	A settlement in romanized text, without context. For example "Nanae, Atsuta, Hekinan", or "Banqiao District".
pii/address/settlement/nocontext/CC	A settlement in CJKVT native script or romanized text, without context. For example "津島市城南区月形町", "板橋區", or "Nanae, Atsuta, Hekinan".
pii/address/settlement/landmark/CC	A settlement landmark. For example "市区町村", "

Entity	Description
	鄉鎮市區", or "City/Ward/Town/Village".
pii/address/region/context/CC	A region, with context. For Taiwan, this is a county (縣 xiàn) or municipality (市 shì). For example "都道府県:神奈川県", 縣市:宜蘭縣", or "Prefecture: Kagoshima".
pii/address/region/nocontext/cjkvt/CC	A region in CJKVT native script, without context. For example "神奈川県", or "宜蘭縣".
pii/address/region/nocontext/latin/CC	A region in romanized text, without context. For example "Kagoshima", or "Yilan County".
pii/address/region/nocontext/CC	A region in CJKVT native script or romanized text, without context. For example "神奈川県", "新北市", or "Kagoshima".
pii/address/region/landmark/CC	A region landmark. For example "都道府県", "縣市", or "Prefecture".
pii/address/postcode/context/CC	A postal code, with context. For example "郵便番号: １２３-４５６７", "Postcode: 1234567", or "郵遞區號106-409".
pii/address/postcode/nocontext/CC	A postal code, without context. For example "１２３-４５６７", or "106-409".
pii/address/postcode/landmark/CC	A postal code landmark. For example "郵便番号", "郵遞區號", or "Postcode".
pii/address/country/context/CC	A country, with context. For example "国:日本", "国:中華民國", or "Country: Japan".
pii/address/country/nocontext/cjkvt/CC	A country in CJKVT native script, without context. For example "日本", or "中華民國".
pii/address/country/nocontext/latin/CC	A country in romanized text, without context. For example "Japan".
pii/address/country/nocontext/CC	A country in CJKVT native script or romanized text, without context. For example "日本" or "Japan".
pii/address/country/landmark/CC	A country landmark. For example "国" or "Country".

date and date_cjkvt (ECR and EJR available)

Entity	Description
pii/date/dob/context/LLL	A date of birth, written numerically or using words. For example "date of birth 1/1/2018", "GEBOORTEDATUM: 01/01/2018"
pii/date/nocontext/LLL	A calendar date, written numerically or using words, without context. For example "01.03.1918", "2018_01_01", "вторник, 30 октомври 2018".
pii/date/dob/landmark/LLL	A date of birth landmark, such as "DOB" or "Fecha de nacimiento".

device_id (ECR and EJR available)

Entity	Description
pii/device_id/ip/nocontext	An IP address, without context.
pii/device_id/imei/nocontext	An IMEI (International Mobile Equipment Identity), without context.
pii/device_id/imeisv/nocontext	An IMEISV (International Mobile Equipment Identity software version) , without context.
pii/device_id/mac_ address/nocontext	A MAC address, without context.
pii/device_id/meid/nocontext	A MEID (Mobile Equipment Identifier), without context.
pii/device_id/iccid/nocontext	An ICCID (Integrated Circuit Card Identifier), without context.
pii/device_id/imsi/nocontext	An IMSI (International Mobile Subscriber Identity), without context.
pii/device_id/msisdn/nocontext	A MSISDN (Mobile Station International Subscriber Directory Number)

driving and driving_cjkvt (ECR and EJR available)

Entity	Description
pii/driving/context/CC	A driving license number with context. For

Entity	Description
	example: "australian automobile association: 103 805 501", or "driver's license: A234567890". This entity matches both the driving license number, and the personal number or driver number, if present. On the standard European driving license, these are fields 5 and 4d.
pii/driving/nocontext/CC	A driving license number, without context.
pii/driving/landmark/CC	A driving license landmark, such as "Driver's license" or "Driving Licence".

health (ECR and EJR available)

Entity	Description
pii/health/ehic/context/CC	An EHIC personal identification number with context. For example "EHIC: UK 1234 5678 " or "TSE: 123456789012".
pii/health/ehic/nocontext/CC	An EHIC personal identification number without context. For example "123456-789A".
pii/health/ehic/landmark/CC	An EHIC landmark, such as "EHIC" or "EHIC PIN".
pii/health/ehic/context/all	An EHIC personal identification number with context, for EU countries and Switzerland.
pii/health/ehic/nocontext/all	An EHIC personal identification number without context, for EU countries and Switzerland.
pii/health/ehic/landmark/all	An EHIC landmark, such as "EHIC" or "EHIC PIN", for EU countries and Switzerland.
pii/health/id/context/au	An Australian Medicard (card) number or Individual Healthcare Identifier (IHI) with context. For example "Medicare Card Number: 3501 80315 1-6".
pii/health/id/context/br	A Brazilian Cartão Nacional de Saúde (CNS, also known as SUS) number with context, for example "CNS: 190129759240018".
pii/health/id/context/ca	A Canadian health insurance (card) number with context. For example "health insurance: 12345-6789", or "assurance-maladie: 12345-6789".
pii/health/id/context/ch	A Swiss health insurance card number with context. For example "Schweizerische Krankenversicherungskarte: 12345678901234567890".

Entity	Description
pii/health/id/context/es	A Spanish health insurance card number with context. For example "CatSalut: ABCD 1 123456 12 1".
pii/health/id/context/fr	A French Carte Vitale number with context. For example "INSEE: 187090100100141".
pii/health/id/context/gb	A British NHS number with context. For example "NHS Number: 943 476 5919".
pii/health/id/context/nz	A New Zealand National Health Index (NHI) number with context. For example "NHI Number: CGC2720".
pii/health/id/context/us	A US health insurance number with context. For example "Medicare ID: 1EG4-TE5-MK72".
pii/heath/id/nocontext/CC	A health number, such as a British NHS number or French Carte Vitale number, without context.
pii/health/id/landmark/CC	A health number landmark, such as "NHS number" or "Medicare ID".

health_cjkvt (ECR and EJR available)

Entity	Description
pii/health/id/context/CC	A heath number with context. For example "記号 ２１７００２３". This entity gives a lower score to matches with more ambiguous landmarks (such as 番号 and "Number"). However, if two matches occur together (for example "記号 21700023 番号 21"), the entity can match both with a higher score.
pii/heath/id/nocontext/CC	A health number, such as a Japanese Health Insurance Card number, without context.
pii/health/id/landmark/CC	A health number landmark, such as "保険者番号" or "Insurer number".

medical_terms.ecr

Entity	Description
pii/medical_terms/LLL	A medical condition or procedure. For example "abdominal hernia". This entity is available for all GDPR languages.

mrtd (ECR and EJR available)

Entity	Description
pii/mrtd/mrp	A machine readable passport. For example "P<GBRUK<SPECIMEN<<ANGELA<ZOE<<<<<<<<<<<<<<<< 5333244280GBR8812049F2509286<<<<<<<<<<<<<<00"
pii/mrtd/mrotd/td1	A machine readable TD1-size travel document. For example "IDD<<T220001293<<<<<<<<<<<<<<< 6408125<2010315D<<<<<<<<<<<<4 MUSTERMANN<<ERIKA<<<<<<<<<<<<"

mrtd_cjkvt (ECR and EJR available)

Entity	Description
pii/mrtd/mrp_cjkvt	A CJKVT machine readable passport line. For example "P<JPN<<<<<<<KEIKO<INOUE<<<<<<<<<<<<<<<<<<<"
pii/mrtd/mrotd/td1_cjkvt	A CJKVT machine readable TD1-size travel document line. For example "KEIKO<<INOUE<<<<<<<<<<<<<<<<<"

name.ecr

Entity	Description
pii/name/CC	A full personal name, in title case or upper case. This entity returns the names in a normalized format, in the form *GIVEN NAME SURNAME*, for example JOHN SMITH. You can turn off normalization by setting `normalize_names=false` in the `name_stoplist.lua` script. You can also turn off score adjustment, by setting `rescore_names=false` in the `name_stoplist.lua` script. This option can improve performance when you do not need the normalization or score refinement.
pii/name/landmark/CC	A full name landmark. For example "name".
pii/name/given_name/context/CC	A given name, with context. For example "Forename: John".
pii/name/given_name/nocontext/CC	A given name, without context. For example "John".

Entity	Description
pii/name/given_name/landmark/CC	A given name landmark. For example "Forename".
pii/name/surname/context/CC	A surname with context. For example "Surname: Smith".
pii/name/surname/nocontext/CC	A surname without context. For example "Smith".
pii/name/surname/landmark/CC	A surname landmark. For example "Surname".
pii/name/pre_title/CC	A title that precedes a name. For example "Ms".
pii/name/post_title/CC	A title that follows a name. For example "Esq".

name_cjkvt.ecr

Entity	Description
pii/name/CC	A full personal name, in romanized text or CJKVT native script. Romanized names can be in title case or upper case, and can be in the order *given name surname* or *surname given name*. CJKVT native script names must be *surname given name*. For Japanese, either form can include honorifics. This entity returns the names in a normalized format, in the form *GIVEN NAME SURNAME*, for example KEIKO NAKAMURA. You can turn off normalization by setting `normalize_names=false` in the `name_stoplist.lua` script. You can also turn off score adjustment, by setting `rescore_names=false` in the `name_stoplist.lua` script. This option can improve performance when you do not need the normalization or score refinement.
pii/name/cjkvt/CC	A full personal name in CJKVT native script. For example "山田恵".
pii/name/latin/CC	A romanized full personal name. For example "Shinzo Abe".
pii/name/landmark/CC	A full name landmark. For example "名前".
pii/name/given_name/context/cjkvt/CC	A given name in CJKVT native script, with context. For example "名前: 直樹 ".
pii/name/given_name/nocontext/cjkvt/CC	A given name in CJKVT native script, without context. For example "直樹 ".
pii/name/given_	A romanized given name, with context. For example "Given

Entity	Description
name/context/latin/CC	Name: Keiko".
pii/name/given_ name/nocontext/latin/CC	A romanized given name, without context. For example "Keiko".
pii/name/given_name/context/CC	A given name in romanized text or CJKVT native script, with context. For example "名前: 直樹 ".
pii/name/given_name/nocontext/CC	A given name in romanized text or CJKVT native script, without context. For example "直樹 ".
pii/name/given_name/landmark/CC	A given name landmark in CJKVT native script. For example: "名前"
pii/name/surname/context/cjkvt/CC	A surname in CJKVT native script, with context. For example "名字: 山田".
pii/name/surname/nocontext/cjkvt/CC	A surname in CJKVT native script, without context. For example "山田".
pii/name/surname/context/latin/CC	A romanized surname, with context. For example "Surname: Nakamura".
pii/name/surname/nocontext/latin/CC	A romanized surname, without context. For example "Nakamura".
pii/name/surname/context/CC	A surname in romanized text or CJKVT native script, with context. For example "名字: 山田".
pii/name/surname/nocontext/CC	A surname in romanized text or CJKVT native script, without context. For example "山田".
pii/name/surname/landmark/jp	A surname landmark in CJKVT native script. For example "名字".
pii/name/pre_title/nocontext/CC	A title that precedes a name in romanized text. For example "Ms".
pii/name/post_title/nocontext/latin/jp	A title that follows a name in romanized text. For example "Esq".
pii/name/post_title/nocontext/cjkvt/jp	A title that follows a name in Japanese script. For example "さん".
pii/name/post_title/nocontext/jp	A title that follows a name in romanized text or Japanese script. For example "Esq" or "さん".

national_id and national_id_cjkvt (ECR and EJR available)

Entity	Description
pii/id/context/CC	A national identity number with context. For information about the supported ID numbers, see Supported National ID Numbers, on page 47. **NOTE:** By default, when there are multiple possible matches, post-processing returns only one country, which is the most efficient option. You can enable ambiguous entity matching by setting `ambiguous_tin_id_entities=true` in the `pii_postprocessing.lua` script. This option returns multiple possible country matches.
pii/id/nocontext/CC	A national identity number without context. For information about the supported ID numbers, see Supported National ID Numbers, on page 47. **NOTE:** By default, when there are multiple possible matches, post-processing returns only one country, which is the most efficient option. You can enable ambiguous entity matching by setting `ambiguous_tin_id_entities=true` in the `pii_postprocessing.lua` script. This option returns multiple possible country matches.
pii/id/landmark/CC	A national identity number landmark, such as "National insurance number" or "Social security number".

nationality and nationality_cjkvt (ECR and EJR available)

Entity	Description
pii/nationality/adj/context/CC	A nationality adjective with context. For example, "Nationality: British".
pii/nationality/adj/nocontext/CC	A nationality adjective without context. For example, "British".
pii/nationality/adj/landmark/CC	A nationality adjective landmark. For example, "Nationality".
pii/nationality/noun/context/CC	A nationality noun with context. For example, "Country: Britain".
pii/nationality/noun/nocontext/CC	A nationality noun without context. For example: "Britain".

Entity	Description
pii/nationality/noun/landmark/CC	A nationality noun landmark. For example, "Country".
pii/nationality/any/context/CC	Any combination of nationality adjective and noun landmark and value. For example, "Country: British", or "Nationality: British".
pii/nationality/any/nocontext/CC	Any nationality adjective or noun. For example, "Britain" or "British".
pii/nationality/any/landmark/CC	Any nationality adjective or landmark. For example, "Nationality" or "Country".

passport and passport_cjkvt (ECR and EJR available)

Entity	Description
pii/passport/context/CC	A passport number with context. For example "Passport number: 533324428", "Passport Number: P4366918", or "italian passaporti AA5275702".
pii/passport/nocontext/CC	A passport number without context. For example "533324428", "C015918", or "14CV28142".
pii/passport/landmark/CC	A passport landmark, such as "Passport" or "Pasaporte". For information about cases where the landmark and passport number do not match or have an ambiguous match, see Ambiguous Entities, on page 83.

postcode and postcode_cjkvt

Entity	Description
pii/postcode/context/CC	A postal code with context. For example "PLZ: 1210", "Poštanski broj: 10000", or "Cod poştal: 235200".
pii/postcode/nocontext/CC	A postal code without context. For example "2700-439 AMADORA", "75018", or "W1B 5TG".
pii/postcode/landmark/CC	A postal code landmark, such as "Postcode" or "Postleitzahl".

telephone.ecr, telephone_cjkvt.ecr and telephone_cjkvt.ejr

Entity	Description
pii/telephone/context/CC	A telephone number with context. For example "Tel: +44 1234 224050", "Telephone: (204)-243-9955", or "numéro de téléphone: +1-902-861-7000". For the `telephone_cjkvt` grammar, numbers can be ASCII or full-width numbers. **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 19. This entity returns the telephone number in the normalized format `+NNNNN`, starting with the country code. For example +12042439955.
pii/telephone/nocontext/CC	A telephone number without context. For example: "(204)-243-9955", "+39 055 326 43 11", or "44 20 7499 9000". For the `telephone_cjkvt` grammar, numbers can be ASCII or full-width numbers. **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 19. This entity returns the telephone number in the normalized format `+NNNNN`, starting with the country code. For example +12042439955.
pii/telephone/landmark/CC	A telephone number landmark, such as "Tel:" or "Telefon No". For the `telephone_cjkvt` grammar, landmarks are available only in CJKVT native script.

tin and tin_cjkvt (ECR and EJR available)

Entity	Description
pii/tin/context/CC	A tax identification number with context. For example "ITIN: 911-92-3333", or "TIN-numre: 101111113". For more examples, see Example Tax Identification Numbers, on page 72. **NOTE:** By default, when there are multiple possible matches, post-processing returns only one country, which is the most efficient option. You can enable ambiguous entity matching by setting `ambiguous_tin_id_entities=true` in the `pii_postprocessing.lua` script. This option returns multiple possible country matches.
pii/tin/nocontext/CC	A tax identification number without context. For example "756.3047.5009.62", or "Z1234567R". For more examples, see Example Tax Identification Numbers, on page 72. **NOTE:** By default, when there are multiple possible matches, post-processing returns only one country, which is the most efficient option. You can enable ambiguous entity matching by setting `ambiguous_tin_id_entities=true` in the `pii_postprocessing.lua` script. This option returns multiple possible country matches.
pii/tin/landmark/CC	A tax identification number landmark, such as "ITIN" or "TIN-numre".

travel (ECR and EJR available)

Entity	Description
pii/travel/context/us	A US passport card number with context. For example "Passport card number: C12345678".
pii/travel/nocontext/us	A US passport card number without context. For example "C12345678".
pii/travel/landmark/us	A US passport card number landmark. For example "Passport card number".

Combined Entities

In addition to the entities described in the Eduction Grammar Reference, on page 20, the IDOL PII Package includes grammar files that contain "combined" entities. These files are named `combined_*.ecr` (or `combined_*_cjkvt.ecr` for Japan) and the entities match addresses, dates, driving license numbers, and so on, from multiple countries.

- The entities that end in `/all` match data for any supported non-CJKVT country or language.
- The entities that end in `/all_cjkvt` match data for any supported CJKVT country.
- The entities that end in `/gdpr` match data for any supported country or language subject to GDPR.

For example:

- Using `pii/address/all` from `combined_address.ecr` matches a postal address from any non-CJKVT country. This is similar to using the `address.ecr` grammar file and extracting `pii/address/??`.
- Using `pii/address/gdpr` from `combined_address.ecr` matches a postal address from any country subject to GDPR. This is similar to using the `address.ecr` grammar file and extracting `gdpr/address/??`.
- Using `pii/date/dob/context/all` from `combined_date.ecr` matches a date of birth written numerically or using words in any language. This is similar to using the `date.ecr` grammar file and extracting `pii/date/dob/context/???`.

The combined (`/all`, `/all_cjkvt` and `/gdpr`) entities provide a significant improvement in processing speed when you extract matches for all countries or languages.

You must run the script `pii_postprocessing.lua` as a post-processing task (see Configure Post Processing, on page 16). Running the script ensures that the entity names returned by Eduction 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 `pii/address/gb`, and not `pii/address/all`.

The combined grammar files might produce fewer matches, because (by default) only a single match is returned in cases where the same characters in the input text would match multiple countries or languages.

> **TIP:** If you need all matches, you can turn on the `AllowMultipleResults` configuration option. This option slows down the matching process because it does not stop after a single match, but is generally still faster than using the individual grammars.

File	Entity
combined_address.ecr	pii/address/all
	pii/address/gdpr
	pii/address/streetlocation/context/all
	pii/address/streetlocation/context/gdpr
	pii/address/city/context/all
	pii/address/city/context/gdpr
	pii/address/country/context/all
	pii/address/country/context/gdpr
	pii/address/postcode/context/all
	pii/address/postcode/context/gdpr
combined_address_cjkvt.ecr	pii/address/all_cjkvt
	pii/address/address1/context/all_cjkvt
	pii/address/region/context/all_cjkvt
	pii/address/country/context/all_cjkvt
	pii/address/postcode/context/all_cjkvt
combined_date.ecr	pii/date/dob/context/all
	pii/date/dob/landmark/all
	pii/date/dob/context/gdpr
	pii/date/dob/landmark/gdpr
	pii/date/nocontext/all
	pii/date/nocontext/gdpr
combined_date_cjkvt.ecr	pii/date/dob/context/all_cjkvt
	pii/date/dob/landmark/all_cjkvt
	pii/date/nocontext/all_cjkvt

File	Entity
combined_driving.ecr	pii/driving/context/all
	pii/driving/nocontext/all
	pii/driving/landmark/all
	pii/driving/context/gdpr
	pii/driving/nocontext/gdpr
	pii/driving/landmark/gdpr
combined_driving_cjkvt.ecr	pii/driving/context/all_cjkvt
	pii/driving/nocontext/all_cjkvt
	pii/driving/landmark/all_cjkvt
combined_health.ecr	pii/health/ehic/context/gdpr
	pii/health/ehic/nocontext/gdpr
	pii/health/ehic/landmark/gdpr
	pii/health/id/context/all
	pii/health/id/nocontext/all
	pii/health/id/landmark/all
	pii/health/id/context/gdpr
	pii/health/id/nocontext/gdpr
	pii/health/id/landmark/gdpr
combined_health_cjkvt.ecr	pii/health/id/context/all_cjkvt
	pii/health/id/nocontext/all_cjkvt
	pii/health/id/landmark/all_cjkvt
combined_name.ecr	pii/name/all
	pii/name/gdpr
combined_name_cjkvt.ecr	pii/name/all_cjkvt
	pii/name/latin/all_cjkvt
	pii/name/cjkvt/all_cjkvt

File	Entity
combined_national_id.ecr	pii/id/context/all
	pii/id/nocontext/all
	pii/id/landmark/all
	pii/id/context/gdpr
	pii/id/nocontext/gdpr
	pii/id/landmark/gdpr
combined_national_id_cjkvt.ecr	pii/id/context/all_cjkvt
	pii/id/nocontext/all_cjkvt
	pii/id/landmark/all_cjkvt
combined_passport.ecr	pii/passport/context/all
	pii/passport/nocontext/all
	pii/passport/landmark/all
	pii/passport/context/gdpr
	pii/passport/nocontext/gdpr
	pii/passport/landmark/gdpr
combined_passport_cjkvt.ecr	pii/passport/context/all_cjkvt
	pii/passport/nocontext/all_cjkvt
	pii/passport/landmark/all_cjkvt
combined_postcode.ecr	pii/postcode/context/all
	pii/postcode/nocontext/all
	pii/postcode/landmark/all
	pii/postcode/context/gdpr
	pii/postcode/nocontext/gdpr
	pii/postcode/landmark/gdpr
combined_postcode_cjkvt.ecr	pii/postcode/context/all_cjkvt
	pii/postcode/nocontext/all_cjkvt
	pii/postcode/landmark/all_cjkvt

File	Entity
combined_telephone.ecr	pii/telephone/context/all
	pii/telephone/nocontext/all
	pii/telephone/landmark/all
	pii/telephone/context/gdpr
	pii/telephone/nocontext/gdpr
	pii/telephone/landmark/gdpr
combined_telephone_cjkvt.ecr	pii/telephone/context/all_cjkvt
	pii/telephone/nocontext/all_cjkvt
	pii/telephone/landmark/all_cjkvt
combined_tin.ecr	pii/tin/context/all
	pii/tin/nocontext/all
	pii/tin/landmark/all
	pii/tin/context/gdpr
	pii/tin/nocontext/gdpr
	pii/tin/landmark/gdpr
combined_tin_cjkvt.ecr	pii/tin/context/all_cjkvt
	pii/tin/nocontext/all_cjkvt
	pii/tin/landmark/all_cjkvt

Components

Some of the PII entities extract *components* as well as whole matches. Components are parts of a match that can provide useful information. For example, the entity for an address can return components for the street and postcode.

> **NOTE:** The post-processing script requires components. The default PII configuration file enables components. If you use a custom configuration file you must set the `EnableComponents` parameter to `True` to return components.

The following tables list the components available for particular entities.

address.ecr - address entity

Component Name	Notes
APARTMENT	
APARTMENT_BUILDING	tr and za
APARTMENT_PREFIX	ca only
BLOCK	br only
CITY	
COUNTRY	
COUNTY	gb only
FLOOR	ca, us, tr, and za
INTERSECTING_STREET	za only
MUNICIPALITY	za only
NUMBER	
PO_BOX	
POSTCODE	
POST_OFFICE	ca, it, lt, no, nz, pt, and za
PRIVATE_BAG_AGENCY	za only
PRIVATE_BAG_BOX_NUMBER	za only
PRIVATE_BAG_NUMBER	za only
PROVINCE	ca only
QUADRANT	br only
REGION	br and za
SECTOR	br only
SITE_ID	za only
SITE_TYPE	za only
STATE	us and au
STREET	
SUBURB	za only

address.ecr - address entity, continued

Component Name	Notes
TOWN	za only
TOWNSHIP	za only
VILLAGE_NAME	za only

The following examples demonstrate the use of these components.

- SHS Quadra 4 Bloco D, 70314-000, Brasília, Brazil

  ```
  SECTOR: SHS
  QUADRANT: QUADRA 4
  BLOCK: BLOCO D
  CITY: BRASÍLIA
  POSTCODE: 70314000
  COUNTRY: BRAZIL
  ```

- Avenida João Jorge, 112, apto. 31, Campinas - SP, 13035-680, Brazil

  ```
  NUMBER: 112
  APARTMENT: APT 31
  STREET: AVENIDA JOÃO JORGE
  CITY: CAMPINAS
  REGION: SP
  POSTCODE: 13035680
  COUNTRY: BRAZIL
  ```

- 2-400 Steeprock Dr Toronto M3J 2X1, Canada

  ```
  APARTMENT_PREFIX: 2-
  NUMBER: 400
  STREET: STEEPROCK DR
  CITY: TORONTO
  POSTCODE: M3J2X1
  COUNTRY: CANADA
  ```

- Çınar mahallesi. Basefendi sok. Küçükyalı Maltepe Asli apt. No:15 Kat 3, Maltepe, 34841 Istanbul, Turkey

  ```
  APARTMENT_BUILDING: KÜÇÜKYALI MALTEPE ASLI APT
  STREET: ÇINAR MAHALLESI BASEFENDI SOK
  NUMBER: 15
  FLOOR: KAT:3
  POSTCODE: 34841
  CITY: ISTANBUL
  COUNTRY: TURKEY
  ```

- PO Box 34, DULUTH MN 55803-0034

```
PO_BOX: 34
CITY: DULUTH
STATE: MN
POSTCODE: 558030034
```

- PO BOX 90242, AUCKLAND MAIL CENTRE, AUCKLAND, 1142

```
PO_BOX: 90242
POST_OFFICE: AUCKLAND MAIL CENTRE
CITY: AUCKLAND
POSTCODE: 1142
```

- 70 Park Street West, Hatfield, City of Tshwane, Gauteng

```
NUMBER: 70
STREET: PARK STREET WEST
SUBURB: HATFIELD
MUNICIPALITY: CITY OF TSHWANE
REGION: GAUTENG
```

- Commission House, corner of Church Avenue and Hill Streets, Pretoria

```
APARTMENT_BUILDING: COMMISSION HOUSE
INTERSECTING_STREET: CHURCH AVENUE
STREET: HILL STREET
CITY: PRETORIA
```

- Farm 938 Rietfontein 341-JR

```
SITE_TYPE: FARM
NUMBER: 938
SUBURB: RIETFONTEIN
SITE_ID: 341-JR
```

- Postnet suite 8839, Private Bag X09, WELTEVREDENPARK, 1715

```
PRIVATE_BAG_AGENCY: POSTNET SUITE
PRIVATE_BAG_BOX_NUMBER: 8839
PRIVATE_BAG_NUMBER: X09
SUBURB: WELTEVREDENPARK
POSTCODE: 1715
```

- 110101 Corana, Umtata, 5100

```
NUMBER: 110101
VILLAGE_NAME: CORANA
TOWN: UMTATA
POSTCODE: 5100
```

- SA Post Office Ltd., PO Box 10 000, PROTEA PARK, 0305

```
POST_OFFICE: SA POST OFFICE LTD
PO_BOX: 10000
TOWNSHIP: PROTEA PARK
POSTCODE: 0305
```

- 123 London Street, Cambridge, Cambridgeshire, CB42AA

  ```
  NUMBER: 123
  STREET: LONDON STREET
  CITY: CAMBRIDGE
  COUNTY: CAMBRIDGESHIRE
  POSTCODE: CB42AA
  ```

- 100 Boulevard Alexis-Nihon, Montreal (Quebec) H4M 2N7

  ```
  NUMBER: 100
  STREET: BOULEVARD ALEXIS NIHON
  CITY: MONTREAL
  PROVINCE: QUEBEC
  POSTCODE: H4M 2N7
  ```

address_cjkvt.ecr - address entity

Component Name	Notes
ALLEY	tw only
APARTMENT	
BLOCK	jp only
BU_BLOCK	jp only
COUNTRY	
CITY	
CITY_BLOCK	jp only
CITY_BLOCK_DIRECTION	jp only
CITY_BLOCK_LANDMARK	jp only
CITY_BLOCK_NUMBER	jp only
DISTRICT	tw only
DISTRICT_DIRECTION	jp only
DISTRICT_NAME	jp only
DISTRICT_NUMBER	jp only
DISTRICT_NUMBER_JIKKAN	jp only
FLOOR	tw only
LANE	tw only

address_cjkvt.ecr - address entity, continued

Component Name	Notes
NEIGHBORHOOD	tw only
NUMBER	
PO_BOX	
POSTCODE	
POST_OFFICE	
REGION	
RURAL_AREA	jp only
RURAL_SUBUNIT	jp only
SECTION	tw only
STREET	tw only
TOWN	jp only
TOWNSHIP	tw only
VILLAGE	
WARD	jp only

The following examples demonstrate the use of these components:

- 日本、〒123-4567神奈川県津島市城南区月形町八重洲四丁目1番2-34号

  ```
  COUNTRY: 日本
  POSTCODE: 1234567
  REGION: 神奈川県
  CITY: 津島市
  WARD: 城南区
  TOWN: 月形町
  DISTRICT_NAME: 八重洲
  DISTRICT_NUMBER: 四
  BLOCK: 1
  NUMBER: 2
  APARTMENT: 34
  ```

- 京都府東筑摩郡音威子府村大字滝沢字住吉6番地の6-22

  ```
  REGION: 京都府
  RURAL_AREA: 東筑摩郡
  VILLAGE: 音威子府村
  RURAL_SUBUNIT: 大字滝沢字住吉
  ```

BLOCK: 6
NUMBER: 6
APARTMENT: 22

- 札幌市中央区南二十二条西十六丁目 2番地

  CITY: 札幌市
  WARD: 中央区
  CITY_BLOCK_DIRECTION: 南
  CITY_BLOCK_NUMBER: 二十二
  DISTRICT_DIRECTION: 西
  DISTRICT_NUMBER: 十六
  BLOCK: 2

- 北海道札幌市厚別区厚別中央二条十丁目 3番 6号

  REGION: 北海道
  CITY: 札幌市
  WARD: 厚別区
  CITY_BLOCK_LANDMARK: 厚別中央
  CITY_BLOCK_NUMBER: 二
  DISTRICT_NUMBER: 十
  BLOCK: 3
  NUMBER: 6

- 札幌市中央区大通西二十丁目 2番 18号

  CITY: 札幌市
  WARD: 中央区
  CITY_BLOCK: 大通
  DISTRICT_DIRECTION: 西
  DISTRICT_NUMBER: 二十
  BLOCK: 2
  NUMBER: 18

- 新潟県新潟市江南区楚川甲180-1

  REGION: 新潟県
  CITY: 新潟市
  WARD: 江南区
  DISTRICT_NAME: 楚川
  DISTRICT_NUMBER_JIKKAN: 甲
  BLOCK: 180
  NUMBER: 1

- 鹿島郡中能登町良川ト部 32番地の5

  RURAL_AREA: 鹿島郡
  TOWN: 中能登町
  DISTRICT_NAME: 良川
  BU_BLOCK: ト
  BLOCK: 32
  NUMBER: 5

- 〒963-8642郡山郵便局私書箱12号

  ```
  POSTCODE: 9638642
  POST_OFFICE: 郡山郵便局
  PO_BOX: 12
  ```

- 10603台北市大安區金山南路2段55號

  ```
  POSTCODE: 10603
  REGION: 台北市
  DISTRICT: 大安區
  STREET: 金山南路
  SECTION: 2
  NUMBER: 55
  ```

- 臺東縣關東河鄉都蘭村42鄰431之2號

  ```
  REGION: 臺東縣
  TOWNSHIP: 關東河鄉
  VILLAGE: 都蘭村
  NEIGHBORHOOD: 42
  NUMBER: 431
  APARTMENT: 2
  ```

- 臺北市大安區延吉街70巷5弄6號1樓

  ```
  REGION: 臺北市
  DISTRICT: 大安區
  STREET: 延吉街
  LANE: 70
  ALLEY: 5
  NUMBER: 6
  FLOOR: 1
  ```

Supported National ID Numbers

The following table lists the national ID numbers that are supported by the `pii/id/context/CC` and `pii/id/nocontext/CC` Eduction entities in the `national_id` grammar (ECR and EJR).

Country	Supported national identity numbers	Example context	Example Match
Argentina	Documento Nacional de Identidad	Registro Nacional de las Persona	22691903
Australia	ImmiCard	ImmiCard	AMS123456
Austria	SSN (social security number) CRR	ASVG	1788011550
Bahrain	ID card	ID card	520202236
Belgium	NRN (numéro de registre national)	numéro national	85 07 30 033 28
Bulgaria	EGN (Uniform Civil Number)	EGN	8032056031
Brazil	Registro de Identidade Civil (RIC) Registro Geral (RG)	RIC Registro Geral	12345678901 56.843.539-4
Canada	Canadian national ID (Social Insurance Number)	numéro d'assurance sociale	159749357
Colombia	'Cédula de Ciudadanía (Número Único de Identificación Personal)	Cédula de Ciudadanía	1.077.650.154
Croatia	OIB (Osobni identifikacijski broj)	OIB	79423753532
Cyprus	Identity card number	Αριθμός ταυτότητας	3861811-2
Czech republic	Rodné číslo	rodné číslo	7360285163
Denmark	CPR	legitimation	011118-0001
Dominican Republic	cédula	Cédula de Identidad y Electoral	001-1490565-6
Egypt	Personal Verification Card	بطاقة تحقيق شخصية:	29501023201952

Estonia	IK (isikukood)	Isikukood	37605030299
Finland	Henkilötunnus (Personal identity code)	Henkilötunnus	311280-888Y
France	INSEE code	Code INSEE	187090100100141
Germany	National ID serial number	Personalausweis	T22000129
Greece	National ID card AMKA (social security number)	AMKA	13121199999
Hungary	Personal Identification Number ID card number	Nemzeti személyazonosító jel	58709189997
Iceland	Kennitala	Kennitala	1809872079
India	Indian Permanent Account Number (PAN) Aadhaar	आधार	2094 7051 9541
Indonesia	KTP/NIK	KTP	3203012503770011
Ireland	PPSN (personal public service number)	PPSN	1234567TW
Italy	Codice Fiscale	codice fiscale	RSS MRA 74D22 A001Q
Latvia	Personas kods	personas kods	121282-11212
Liechtenstein	Identitätskarte	Personalausweis	ID98754015
Lithuania	Asmens Kodas	asmens kodas	38409152012
Luxembourg	National ID card number Identity card number	Steuernummer	1893120105732
Malaysia	Malaysian National Registration Identity Card number	MyKad	510317-13-5131
Malta	ID card number	ID card	9999999M
Mexico	NSS CURP	CURP	HEGG560427MVZRRL04
Netherlands	BSN	legitimatiebewijs	269740533
Norway	Fødselsnummer	Fødselsnummer	18098749914

	D-nummer H-nummer FH-nummer		
Pakistan	NIC/CNIC	CNIC	61101-9063070-1
Poland	PESEL	PESEL	44051401359
Portugal	Número de identificação civil Cartão de cidadão number Número de Identificação de Segurança Social	NIC	118666070
Romania	Cod Numeric Personal	CNP	1800101221144
Russia	INN	Идентификационный номер налогоплательщика	233502835860
Slovakia	Rodné číslo ID card number	rodné číslo	7360285163
Slovenia	Enotna matična številka občana	EMŠO	1809987504991
South Africa	National ID	Inombolo yesazisi	8001015009087
Spain	DNI NIE	DNI	00000000T
Sweden	Personnummer Samordningsnummer	personnummer	870918-9990
Switzerland	AHV number	AHV-Nr	756.1234.5678.97
Turkey	Turkish Identification Number	türkiye cumhuriyeti kimlik numarası	98768109974
UK	National Insurance Number	National Insurance Number	AB 12 34 56 A
United Arab Emirates	United Arab Emirates identity card number	بطاقة الهوية	784-2001-1234566-1
US	US Social Security Number	SSN	111-22-3333

The following table lists the national ID numbers that are supported by the `pii/id/context/CC` and `pii/id/nocontext/CC` Eduction entities in the `national_id_cjkvt` grammar (ECR and EJR).

Country	Supported national identity numbers	Example context	Example Match
China	Chinese Resident Identity Card number	身份证	460032197910193621
Hong Kong	Hong Kong Identity Card number	HKID	Z683365(5)
Japan	My Number	マイナンバー	654327654322
Singapore	Singapore National Registration Identity Card number	NRIC	S7908207D
Taiwan	Taiwan nation identification number	身份證字號	A123456789
Thailand	Thai identity card number	บัตรประจำตัวประชาชนไทย	4-8547-01245-28-9

Eduction Grammar Examples

The following sections contain examples for each of the IDOL PII Package grammars and entities for each supported language.

Example Addresses

The following table lists example landmarks and values for the address grammar (`address.ecr`) for each supported country.

Type	Country	Landmark	Example
pii/address	at		Schlosshoferstrasse 20,1210 Vienna
pii/address	au		100 Flushcombe Road,Albury 2148
pii/address	be		Rue Gregoire Soupart 14,6200,Chatelet
pii/address	bg		Slavyanska 29,1000 SOFIA
pii/address	br		Avenida João Jorge,112,apto. 31,Campinas - SP,13035-680
pii/address	ca		100 Boulevard Alexis-Nihon,Suite 310,Montreal,H4M 2N7,Canada
pii/address	ch		Bahnhofstrasse 4a/8,8001 Zurich
pii/address	cy		Diagorou 29,2097 NICOSIA
pii/address	cz		AUGUSTINOVA 2068,148 00,PRAGUE
pii/address	de		Hostatostrasse 16,65929,Frankfurt

pii/address	dk		Hamletsgade 4,2200,København
pii/address	ee		Vabaduse Väljak 7,15199 TALLINN
pii/address	es		Avenida Juan Xxiii 20,41006,Sevilla
pii/address	fi		Insinoeoerinkatu 27,33720 Tampere
pii/address	fr		6 RUE CHRISTIANI,75018,PARIS
pii/address	gb		162-168 Regent Street,London,W1B 5TG
pii/address	gr		Vithinias 17,54453 THESSALONIKI
pii/address	hr		Krapinska 45,10000 ZAGREB
pii/address	hu		Pesti Barnabas 4,1052 Budapest
pii/address	ie		99 Barrack Street,Cork C12 F9J4
pii/address	is		Huldugil 19,603 Akureyri
pii/address	it		Viale Risorgimento 55,42100 Reggio Emilia
pii/address	li		Palduinstrasse 33,9496 Balzers
pii/address	lt		Gedimino pr. 16,01103 Vilnius
pii/address	lu		42 rue de la Vallée,2661 LUXEMBOURG
pii/address	lv		Ratslaukums 1,RIGA 1050
pii/address	mt		254 Republic Street,VALLETTA VLT 1114
pii/address	nl		SCHIEDAMSEWEG 199B,3026AM ROTTERDAM
pii/address	no		KONGENS GATE 23,3717,SKIEN
pii/address	nz		100 Mount Eden Road,Mount Eden,Auckland 5022
pii/address	pl		Zeganska 2,04-713,Warszawa
pii/address	pt		GONCALVES RAMOS 104 A,2700-439 AMADORA,LISBON
pii/address	ro		Calea Bucuresti 24,235200,Caracal
pii/address	se		Goetaplatsen 9,41134 Stockholm
pii/address	si		Funtkova 46,1000 LJUBLJANA
pii/address	sk		Sibirska 48,83102 BRATISLAVA
pii/address	tr		Kucukayasofya Mah. Donus SOKAK Uyar Apt. No:3 34000 Istanbul,Turkey

pii/address	us		455 Larkspur Dr. Apt 23,L.A.,92808
pii/address	za		70 Park Street West, Hatfield, City of Tshwane, Gauteng

The following table lists example landmarks and values for the CJKVT address grammar (`address_cjkvt.ecr` or `address_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/address	jp		日本、〒123-4567神奈川県津島市城南区月形町八重洲四丁目 1番 2-34号 1-2-34, Yaesu 4-Chome, Nanae, Atsuta, Hekinan, Kagoshima, 123-4567, Japan
pii/address	tw		10603台北市大安區金山南路 2段 55號 No.55, Sec. 2, Jinshan S. Rd., Daan Dist., Taipei City 10603

Example Dates

The following table lists example landmarks and values for the date grammar (`date.ecr` or `date.ejr`) for each supported country.

Type	Language	Landmark	Example
pii/date/dob	afr	GEBOORTEDATUM	22/07/2020
pii/date/dob	bul	Рожден ден	01/01/2018
pii/date/dob	cat	data de naixement	01/01/2018
pii/date/dob	cze	datum narození	01/01/2018
pii/date/dob	dan	fødselsdato	01/01/2018
pii/date/dob	dut	geboortedatum	01/01/2018
pii/date/dob	eng	DOB	01/01/2018
pii/date/dob	est	sünniaeg	01/01/2018
pii/date/dob	fin	syntymäaika	01/01/2018
pii/date/dob	fre	date de naissance	01/01/2018
pii/date/dob	ger	Geburtsdatum	01/01/2018
pii/date/dob	gle	Dáta breithe	01/01/2018
pii/date/dob	gre	γέννηση	01/01/2018

pii/date/dob	hrv	datum rođenja	01/01/2018
pii/date/dob	hun	született	01/01/2018
pii/date/dob	ice	FÆÐINGARDAGUR	01/01/2018
pii/date/dob	ita	DATA DI NASCITA	01/01/2018
pii/date/dob	lav	DZIMŠANAS DATUMS	01/01/2018
pii/date/dob	lit	GIMIMO DATA	01/01/2018
pii/date/dob	mlt	DATA TAT-TWELID	01/01/2018
pii/date/dob	nor	FØDT	01/01/2018
pii/date/dob	nso	BELEGWE	22/07/2020
pii/date/dob	pol	DATA URODZENIA	01/01/2018
pii/date/dob	por	DATA DE NASCIMENTO	01/01/2018
pii/date/dob	roh	NAT	01-01-2018
pii/date/dob	rum	DATA NAȘTERII	01/01/2018
pii/date/dob	slo	DÁTUM NARODENIA	01/01/2018
pii/date/dob	slv	DATUM ROJSTVA	01/01/2018
pii/date/dob	spa	FECHA DE NACIÓ	01/01/2018
pii/date/dob	ssw	KUTALWA	22/07/2020
pii/date/dob	swe	FÖDELSEDAG	01/01/2018
pii/date/dob	tso	SIKU ROTSWARIWA	22/07/2020
pii/date/dob	tur	DOĞUM YILI	01/01/2018
pii/date/dob	ven	BEBWA	22/07/2020
pii/date/dob	xho	UMHLA WOKUZALWA	22/07/2020
pii/date/dob	zul	USUKU LOKUZALWA	22/07/2020
pii/date	afr		22 Julie 2020
pii/date	bul		30.10.18
pii/date	cat		30/10/18
pii/date	cze		30.10.18
pii/date	dan		30/10/2018

pii/date	dut		30-10-18
pii/date	eng		01/Jan/2018
pii/date	est		30.10.18
pii/date	fin		30.10.2018
pii/date	fre		30/10/2018
pii/date	ger		30.10.18
pii/date	gle		30/10/2018
pii/date	gre		30/10/18
pii/date	hrv		30. 10. 2018
pii/date	hun		2018. 10. 30.
pii/date	ice		30.10.2018
pii/date	ita		30/10/18
pii/date	lav		30.10.18
pii/date	lit		2018-10-30
pii/date	mlt		30/10/2018
pii/date	nor		30.10.2018
pii/date	nso		22 Mosegamanye 2020
pii/date	pol		30.10.2018
pii/date	por		30/10/2018
pii/date	roh		30-10-18
pii/date	rum		30.10.2018
pii/date	slo		30. 10. 2018
pii/date	slv		30. 10. 18
pii/date	spa		30/10/18
pii/date	ssw		22 Khólwáne 2020
pii/date	swe		2018-10-30
pii/date	tso		22 Mawuwani 2020
pii/date	tur		29.03.2019

pii/date	ven		22 Fulwana 2020
pii/date	xho		22 EyeKhala 2020
pii/date	zul		Julayi 22, 2020

The following table lists example landmarks and values for the CJKVT date grammar (`date_cjkvt.ecr` or `date_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/date/dob	chi	生年	1948年12月6日
pii/date/dob	jpn	生年月日	1948年12月6日
pii/date	chi		1948年12月6日
pii/date	jpn		2020年2月7日 H20/12/31

Example Driving Licenses

The following table lists example landmarks and values for the driving license grammar (`driving.ecr` or `driving.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/driving	at	Fahrausweis	12345678
pii/driving	au	australian automobile association	103 805 501
pii/driving	be	rijbewijs	1234567890
pii/driving	bg	Шофьорска карта	123456789
pii/driving	br	permiso de conducción	123456789
pii/driving	ca	driver's licence	PU-BL-IJ-Q108NA
pii/driving	ch	führerausweis	006551496001
pii/driving	cy	άδεια οδήγησης	123456789012
pii/driving	cz	Řidičák	AB 123456
pii/driving	de	Führerschein	A1234567890
pii/driving	dk	Førerbevis	1234567
pii/driving	ee	juhiluba	AA123456
pii/driving	es	Autorización para la conducción de vehículos	12345678-A

pii/driving	fi	Körkort	123456-123B
pii/driving	fr	Permis véhicule léger	12AB12345
pii/driving	gb	driving licence	ABC99851207A99AB12
pii/driving	gr	άδεια οδήγησης	123456789
pii/driving	hr	Vozačka dozvola	12345678
pii/driving	hu	Vezetői engedély	AB1234567
pii/driving	ie	driver's license	A234567890
pii/driving	is	Ökuskírteini	123456789
pii/driving	it	Licenza di condurre	ABCDEABCD0
pii/driving	li	Klasse M	123456789012
pii/driving	lt	Vairuotojo pažymėjimas	12345678
pii/driving	lu	Führerscheinklasse	123456
pii/driving	lv	braukšanas apliecība	AB123456
pii/driving	nl	rijbewijs	1234567890
pii/driving	no	Førerkort	12 34 123456 1
pii/driving	pl	Prawo jazdy	12345/12/1234
pii/driving	pt	Carta de condução	A-123456 1
pii/driving	ro	Permis de conducere	A12345678B
pii/driving	se	B-körkort	123456-1234A
pii/driving	si	vozniško dovoljenje	123456789
pii/driving	sk	Vodičský preukaz	A1234567
pii/driving	tr	Sürücü belgesi	209573
pii/driving	us	Driving license	012AB3456
pii/driving	za	bestuurslisensie	12345678AB90

The following table lists example landmarks and values for the CJKVT driving grammar (`driving_cjkvt.ecr` or `driving_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/driving	jp	運転免許	901103864930
pii/driving	tw	駕駛執照	A123456789

Example Health Numbers

The following table lists example landmarks and values for the health grammar (`health.ecr` or `health.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/health/id	at	EKVK	1234567890
pii/health/id	au	Medicare Card Number	3501 80315 1-6
pii/health/id	be	EZVK	123456 789 01
pii/health/id	bg	ЕЗОК	1234567890
pii/health/id	br	Cartão Nacional de Saúde	190129759240018
pii/health/id	ca		12345-6789
pii/health/id	ch	Europäische Krankenversicherungskarte	756.1234.5678.97
pii/health/id	ch	Schweizerische Krankenversicherungskarte KVG	12345678901234567890
pii/health/id	cz	Evropský průkaz zdravotního pojištění	1234567890
pii/health/id	de	Europäische Krankenversicherungskarte	A123456789
pii/health/id	dk	EU-sygesikringsbevis	1234567890
pii/health/id	ee	Euroopa ravikindlustuskaart	12345678901
pii/health/id	es	CatSalut	ABCD 1 123456 12 1
pii/health/id	es	TSE	123456789012
pii/health/id	fi	Europeiska sjukvårdskortet	123456-789A
pii/health/id	fr	CEAM	1 23 45 67 890 123 45
pii/health/id	fr	INSEE	187090100100141
pii/health/id	gb	EHIC	UK 1234 5678
pii/health/id	gb	NHS Number	943 476 5919
pii/health/id	gr	Ευρωπαϊκή Κάρτα	1234567

		Ασφάλισης Ασθένειας	
pii/health/id	hr	EKZO	123456789
pii/health/id	hu	EGT	123 456 789
pii/health/id	ie	Carta Árachais Sláinte na hEorpa	1234567A
pii/health/id	is	ES kortið	1234567890
pii/health/id	it	TEAM	SSN-MIN SALUTE-123456
pii/health/id	li	Europäische Krankenversicherungskarte	123456789012A
pii/health/id	lt	Europos sveikatos draudimo kortelė	12345678901
pii/health/id	lv	EVAK	1234 AAA
pii/health/id	nl	Europese gezondheidskaart	123 456 789
pii/health/id	no	Europeisk helsetrygdkort	12345678901
pii/health/id	nz	NHI Number	CGC2720
pii/health/id	pl	EKUZ	12345678901
pii/health/id	pt	CESD	123456789012345678AA
pii/health/id	ro	CNAS	1234567890123
pii/health/id	se	EU-kort	12345678-1234
pii/health/id	si	EU KZZ	123456789
pii/health/id	sk	EPZP	1234567890
pii/health/id	us	Medicare ID	1EG4-TE5-MK72

The following table lists example landmarks and values for the CJKVT health grammar (`health_cjkvt.ecr` or `health_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/health/id	jp	保険者番号	21700023
pii/health/id	tw	健保卡	A123456789

Example National IDs

The following table lists example landmarks and values for the national ID grammar (`national_id.ecr` or `national_id.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/id	at	ASVG	1788011550
pii/id	at	Bereichsspezifische Personen-Kennung	j/NxdRQhp+tNyE9WhHdBSYuy3hA=
pii/id	at	sourcePIN	Qq03dPrgcHsx3G0lKSH6SQ==
pii/id	au	ImmiCard	AMS123456
pii/id	be	numéro national	85 07 30 033 28
pii/id	bg	EGN	8032056031
pii/id	br	RG	56.843.539-4
pii/id	ca	Social Insurance number	159749357
pii/id	ch	AHV-Nr	756.1234.5678.97
pii/id	cy	Αριθμός ταυτότητας	3861811-2
pii/id	cz	rodné číslo	7360285163
pii/id	de	Personalausweis	T22000129
pii/id	dk	legitimation	011118-0001
pii/id	ee	Isikukood	37605030299
pii/id	es	DNI	00000000T
pii/id	fi	Henkilötunnus	311280-888Y
pii/id	fr	Code INSEE	187090100100141
pii/id	gb	National Insurance Number	AB 12 34 56 A
pii/id	gr	AMKA	13121199999
pii/id	hr	OIB	79423753532
pii/id	hu	Nemzeti személyazonosító jel	58709189997
pii/id	ie	PPSN	1234567TW
pii/id	is	Kennitala	1809872079

pii/id	it	codice fiscale	RSS MRA 74D22 A001Q
pii/id	li	Personalausweis	ID98754015
pii/id	lt	asmens kodas	38409152012
pii/id	lu	Steuernummer	1893120105732
pii/id	lv	personas kods	121282-11212
pii/id	mt	ID card	9999999M
pii/id	nl	legitimatiebewijs	269740533
pii/id	no	Fødselsnummer	18098749914
pii/id	pl	PESEL	44051401359
pii/id	pt	NIC	118666070
pii/id	ro	CNP	1800101221144
pii/id	se	personnummer	870918-9990
pii/id	si	EMŠO	1809987504991
pii/id	sk	rodné číslo	7360285163
pii/id	tr	Türkiye Cumhuriyeti Kimlik Numarası	98768109974
pii/id	us	Social Security Number	111-22-3333
pii/id	za	Inombolo yesazisi	8001015009087

The following table lists example landmarks and values for the CJKVT national ID grammar (`national_id_cjkvt.ecr` or `national_id_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/id	jp	マイナンバー	６５４３２７６５４３２２
pii/id	tw	身份證字號	A123456789

Example MRTDs

The following table lists example landmarks and values for the MRTD grammar (`mrtd.ecr`or `mrtd.ejr`).

Type	Country	Landmark	Example
pii/mrtd/mrotd/t			IDD<<T220001293<<<<<<<<<<<<<<

d1			
pii/mrtd/mrp			P<GBROTHER<<ADAM<NORMAN<<<<<<<<<<<<<<<<<<<<<

The following table lists example landmarks and values for the CJKVT MRTD grammar (`mrtd_cjkvt.ecr` or `mrtd_cjkvt.ejr`).

Type	Country	Landmark	Example
pii/mrtd/mrotd/td1_cjkvt			KEIKO<<INOUE<<<<<<<<<<<<<<<<<<
pii/mrtd/mrp_cjkvt			P<JPN<<<<<<<KEIKO<INOUE<<<<<<<<<<<<<<<<<<<<<

Example Name

The following table lists example landmarks and values for the Name grammar (`name.ecr`) for each supported country.

Type	Country	Landmark	Example
pii/name	at		Jürgen Schäfer
pii/name	au		Michael Smith
pii/name	be		Luc Janssen
pii/name	bg		ПЕТЪР Iliev
pii/name	br		Monica Duarte
pii/name	ca		Michael Smith
pii/name	ca		Stéphanie Tremblay
pii/name	ch		Jürgen Conti
pii/name	cy		Kostas Οικονόμου
pii/name	cz		Pavel Petr
pii/name	de		Jürgen Schäfer
pii/name	dk		Søren Jensen
pii/name	ee		Anneli Sild
pii/name	es		Pau Romero

pii/name	fi		Elina Partanen
pii/name	fr		Jerome Fournier
pii/name	gb		Stuart Taylor
pii/name	gr		Kostas Οικονόμου
pii/name	hr		Zoran Lučić
pii/name	hu		Zoltán Fekete
pii/name	ie		Stuart Taylor
pii/name	is		Unnur Stefánsson
pii/name	it		Chiara Conti
pii/name	jp		山田直樹
pii/name	li		Jürgen Schäfer
pii/name	lt		Sandra Butkute
pii/name	lu		Jürgen Fournier
pii/name	lv		Roberts Priede
pii/name	mt		Michelle Cardona
pii/name	nl		Patrick Verstraete
pii/name	no		Arne Thorsen
pii/name	nz		Michael Smith
pii/name	pl		Konrad Tomaszewska
pii/name	pt		Raquel Mota
pii/name	ro		Laurentiu Camelia
pii/name	se		Sara Lindholm
pii/name	si		Darja Likar
pii/name	sk		Katka Oravec
pii/name	tr		Klaus Acar
pii/name	us		Michael Smith
pii/name	za		Pieter Swanepoel

The following table lists example landmarks and values for the CJKVT name grammar (`name_cjkvt.ecr` or `name_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/name	jp		山田直樹
pii/name	tw		劉佳穎

Example Nationalities

The following table lists example landmarks and values for the nationality grammar (`nationality.ecr` or `nationality.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/nationality/adj	at	Nationalität	österreichisch
pii/nationality/adj	au	nationality	Australian
pii/nationality/adj	be	nationaliteit	Belgisch
pii/nationality/adj	bg	националност	български
pii/nationality/adj	br	nacionalidade	Brasileiro
pii/nationality/adj	ca	nationality	Canadian
pii/nationality/adj	ch	Nationalität	Schweizer
pii/nationality/adj	cy	εθνικότητα	κυπριακό
pii/nationality/adj	cz	národnost	čeština
pii/nationality/adj	de	Nationalität	deutsch
pii/nationality/adj	dk	nationalitet	dansk
pii/nationality/adj	ee	Rahvas	Eestlane
pii/nationality/adj	es	nacionalidad	español
pii/nationality/adj	fi	kansallisuus	suomi
pii/nationality/adj	fr	nationalité	française
pii/nationality/adj	gb	nationality	British
pii/nationality/adj	gr	εθνικότητα	ελληνική
pii/nationality/adj	hr	Nacionalnost	hrvatska
pii/nationality/adj	hu	nemzetiség	magyar
pii/nationality/adj	ie	nationality	Irish
pii/nationality/adj	is	þjóðerni	íslensku

pii/nationality/adj	it	nazionalità	italiana
pii/nationality/adj	li	Nationalität	liechtensteinischer
pii/nationality/adj	lt	Tautybė	lietuvis
pii/nationality/adj	lu	Nationalitéit	Lëtzebuerger
pii/nationality/adj	lv	Valstspiederība	latvietis
pii/nationality/adj	mt	nationality	Maltese
pii/nationality/adj	nl	nationaliteit	nederlands
pii/nationality/adj	no	nasjonalitet	norsk
pii/nationality/adj	nz	nationality	New Zealander
pii/nationality/adj	pl	narodowość	polska
pii/nationality/adj	pt	nacionalidade	Português
pii/nationality/adj	ro	naționalitate	română
pii/nationality/adj	se	nationalitet	svenska
pii/nationality/adj	si	narodnost	slovensko
pii/nationality/adj	sk	národnosť	slovenská
pii/nationality/adj	tr	Milliyet	türkçe
pii/nationality/adj	us	nationality	American
pii/nationality/adj	za	nasionaliteit	Republiek van Suid Afrika
pii/nationality/noun	at	Land	Österreich
pii/nationality/noun	au	country	Australia
pii/nationality/noun	be	land	Belgien
pii/nationality/noun	bg	страна	България
pii/nationality/noun	br	país	Brasil
pii/nationality/noun	ca	country	Canada
pii/nationality/noun	ch	Land	Schweiz
pii/nationality/noun	cy	χώρα	Κύπρος
pii/nationality/noun	cz	státní útvar	Česko
pii/nationality/noun	de	Land	Deutschland

pii/nationality/noun	dk	land	Danmark
pii/nationality/noun	ee	maa	Eesti
pii/nationality/noun	es	país	España
pii/nationality/noun	fi	maa	Suomi
pii/nationality/noun	fr	pays	France
pii/nationality/noun	gb	country	Afghanistan
pii/nationality/noun	gr	χώρα	Ελλάδα
pii/nationality/noun	hr	zemlja	Hrvatska
pii/nationality/noun	hu	ország	Magyarország
pii/nationality/noun	ie	country	Ireland
pii/nationality/noun	is	Land	Ísland
pii/nationality/noun	it	Paese	Italia
pii/nationality/noun	li	Land	Liechtenstein
pii/nationality/noun	lt	Šalis	Lietuva
pii/nationality/noun	lu	Land	Luxembourg
pii/nationality/noun	lv	zeme	Latvija
pii/nationality/noun	mt	country	Malta
pii/nationality/noun	nl	land	Nederland
pii/nationality/noun	no	land	Norge
pii/nationality/noun	nz	country	New Zealand
pii/nationality/noun	pl	kraj	Polska
pii/nationality/noun	pt	país	Portugal
pii/nationality/noun	ro	țară	România
pii/nationality/noun	se	land	Sverige
pii/nationality/noun	si	Dežela	Slovenija
pii/nationality/noun	sk	Krajina	Slovensko
pii/nationality/noun	tr	ülke	Türkiye
pii/nationality/noun	us	country	America
pii/nationality/noun	za	land	Republiek van Suid Afrika

The following table lists example landmarks and values for the CJKVT nationality grammar (`nationality_cjkvt.ecr` or `nationality_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/nationality/adj	jp	国籍	日本語
pii/nationality/adj	tw	國籍	台灣人
pii/nationality/noun	jp	国	日本
pii/nationality/noun	tw	國家	臺灣

Example Passport Numbers

The following table lists example landmarks and values for the passport grammar (`passport.ecr` or `passport.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/passport	at	Reisepass	P4366918
pii/passport	au	passport number	M1234567
pii/passport	be	№. du passeport	LA080402
pii/passport	bg	паспорт	383641306
pii/passport	br	número do passaporte	FH254787
pii/passport	ca	passport number	LA123456
pii/passport	ch	numero di passaporto	S5200073
pii/passport	cy	διαβατήριο	C015918
pii/passport	cz	cestovni pas	99009054
pii/passport	de	reisepass	C748TJ1K2
pii/passpo	dk	pas	900010172

rt			
pii/passport	ee	pass	K3295837
pii/passport	es	pasaporte	Q979029
pii/passport	fi	passi	XP8271602
pii/passport	fr	passeport	14CV28142
pii/passport	gb	Passport	925665416
pii/passport	gr	διαβατήριο	AK0219304
pii/passport	hr	putovnica	197520213
pii/passport	hu	utlevel	BB0979829
pii/passport	ie	passport	XN5004216
pii/passport	is	vegabref	A1197783
pii/passport	it	passaporto	AA5275702
pii/passport	li	reisepass	R00536
pii/passport	lt	pasas	20200997
pii/passport	lu	passeport	JC4E7L2H
pii/passport	lv	pase	LZ3263061
pii/passport	mrp		P<GBROTHER<<ADAM<NORMAN<<<<<<<<<<<<<<<<<<<<<
pii/passport	mt	passport	1281376

pii/passport	nl	paspoort	NUL8H33B6
pii/passport	no	innenrikspass	25055689
pii/passport	nz	passport number	LA123456
pii/passport	pl	paszport	ZS 8000038
pii/passport	pt	PASSAPORTE	H045489
pii/passport	ro	PASAPORT	052763786
pii/passport	se	PASS	62366643
pii/passport	si	POTNI LIST	PT0000005
pii/passport	sk	CESTOVNY PAS	XB7891988
pii/passport	tr	PASAPORT	U12345678
pii/passport	us	passport #	123456789
pii/passport	za	ipasipoti	439405987

The following table lists example landmarks and values for the CJKVT passport grammar (`passport_cjkvt.ecr` or `passport_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/passport	jp	パスポート	XS１２３４５６７
pii/passport	tw	护照	378945612

Example Postcodes

The following table lists example landmarks and values for the postcode grammar (`postcode.ecr` or `postcode.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/postcode	at	PLZ	1210
pii/postcode	au	POSTCODE	1140
pii/postcode	be	code postal	6200
pii/postcode	bg	Пощиээки код	1000
pii/postcode	br	CEP	13010
pii/postcode	ca	Postcode	K2K 3C9
pii/postcode	ch	PLZ	1003
pii/postcode	cy	Ταχυδρομικός κωδικός	2097
pii/postcode	cz	směrovací číslo	148 00
pii/postcode	de	Postleitzahl	65929
pii/postcode	dk	postkode	2200
pii/postcode	ee	Sihtnumber	15199
pii/postcode	es	código postal	41006
pii/postcode	fi	Postkoder	33720
pii/postcode	fr	code postal	75018
pii/postcode	gb	côd post	W1B 5TG
pii/postcode	gr	Ταχυδρομικος κωδικος	54453
pii/postcode	hr	Poštanski broj	10000
pii/postcode	hu	iranyitoszam	1052
pii/postcode	ie	Eirchod	C12 F9J4
pii/postcode	is	Postnumer	603
pii/postcode	it	CAP	42100
pii/postcode	li	Postleitzahl	9496
pii/postcode	lt	Pasto kodas	01103
pii/postcode	lu	Postalisch	2661
pii/postcode	lv	Pasta indekss	1050
pii/postcode	mt	postal code	VLT 1114

pii/postcode	nl	postnummer	3026AM
pii/postcode	no	postkode	3717
pii/postcode	nz	POSTCODE	1111
pii/postcode	pl	kod pocztowy	04-713
pii/postcode	pt	Codigo postal	2700-439 AMADORA
pii/postcode	ro	Cod poştal	235200
pii/postcode	se	POSTKODER	41134
pii/postcode	si	Poštna številka	1000
pii/postcode	sk	PSČ	83102
pii/postcode	tr	POSTA KODU	31900
pii/postcode	us	ZIP	94070
pii/postcode	za	Ikhodi yeposi	1234

The following table lists example landmarks and values for the CJKVT postcode grammar (`postcode_cjkvt.ecr` or `postcode_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/postcode	jp	〒	123-4567
pii/postcode	tw	郵遞區號	106-409

Example Telephone Numbers

The following table lists example landmarks and values for the telephone grammar (`telephone.ecr`) for each supported country.

Type	Country	Landmark	Example
pii/telephone	at	Telefon	+43 (1) 716130
pii/telephone	au	TELEPHONE	+61-3-9825-2300
pii/telephone	be	telefoon	+32 2 287 62 11
pii/telephone	bg	Телефон	(+359) 2 933 9222
pii/telephone	br	TELEFONE	+55 11 2345 5678
pii/telephone	ca	TELEPHONE	(204)-243-9955

pii/telephone	ch	telefono	021 123 45 67
pii/telephone	cy	Τηλέφωνο	+357 22 861100
pii/telephone	cz	telefonní číslo	+420 257 40 2111
pii/telephone	de	Funktelefon	+49 (0) 30 20457 0
pii/telephone	dk	mobiltelefon	+45 35 44 52 00
pii/telephone	ee	mobiiltelefon	+372 6674 700
pii/telephone	es	teléfono móvil	+34 91 714 6300
pii/telephone	fi	matkapuhelin	+358 (0) 9 2286 5100
pii/telephone	fr	Numero de telephone	+33 (0)1 44 51 31 00
pii/telephone	gb	tel:	020 7008 1500
pii/telephone	gr	Κινητό	+30 210 7272 600
pii/telephone	hr	mobilni telefon	+385 1 6009 100
pii/telephone	hu	telefonszám	+36 (1) 266 2888
pii/telephone	ie	cellphone	+353 (0) 1205 3700
pii/telephone	is	farsími	+354 550 5100
pii/telephone	it	telefonino	+39 06 4220 0001
pii/telephone	li	Natel	+423 399 44 44
pii/telephone	lt	Telefonas	+370 5 246 2900
pii/telephone	lu	Handy	(+352) 22 98 64
pii/telephone	lv	mobilais telefons	00 371 67774700
pii/telephone	mt	phone	+356 2323 0000
pii/telephone	nl	mobiele telefoon	+31 (0)70 4270 427
pii/telephone	no	telefonnummer	+47 2313 2700
pii/telephone	nz	TELEPHONE	+64-3-234-5678
pii/telephone	pl	Numer abonenta	+48 22 311 00 00
pii/telephone	pt	Número telefônico	+351 21 392 4000
pii/telephone	ro	telefon	+40 (21) 201 7200
pii/telephone	se	mobiltelefon	+46 (0) 8 671 30 00

pii/telephone	si	prenosni telefon	+386 1 200 39 10
pii/telephone	sk	Telefónne číslo	+421 2 5998 20 00
pii/telephone	tr	contach phone	+90 (0)312 290 3390
pii/telephone	us	TELEPHONE	(415)-243-9955
pii/telephone	za	umnxeba	041 504 1951

The following table lists example landmarks and values for the CJKVT telephone grammar (`telephone_cjkvt.ecr` or `telephone_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/telephone	jp	電話機	+８１(３) ３２２４ ５０００
pii/telephone	tw	电话	0425 60 9230

Example Tax Identification Numbers

The following table lists example landmarks and values for the tax identification number grammar (`tin.ecr` or `tin.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/tin	at	Steuer-Identifikationsnummer	931736581
pii/tin	au	Tax File Number	123456782
pii/tin	be	Numéro d'identification fiscale	00012511119
pii/tin	bg	Данъчен идентификационен номер	7501010010
pii/tin	br	CNPJ	90.386.859/0002-54
pii/tin	br	CPF	29594421134
pii/tin	ca	Tax Identification Number	159749357
pii/tin	ch	AHV-Nr	756.3047.5009.62
pii/tin	cy	ΑΦΜ	00123123T
pii/tin	cz	Daňové identifikační číslo	7360285163
pii/tin	de	Steuer-Identifikationsnummer	26954371827

pii/tin	dk	TIN-numre	0101111113
pii/tin	ee	Maksukohustuslase number	32708101201
pii/tin	es	Número de identificación fiscal	Z1234567R
pii/tin	fi	Skatteregistreringsnummer	131052-308T
pii/tin	fr	NIF	3023217600053
pii/tin	gb	Taxpayer Identification Number	12345 12345K
pii/tin	gr	Αριθμός Μητρώου Φορολογούμενου	123456789
pii/tin	hr	Identifikacijski porezni broj	94577403194
pii/tin	hu	Adóazonosító szám	8071592153
pii/tin	ie	Uimhir Aitheantais Cánach	1234567TW
pii/tin	is	Skattaupplýsingar	1809872079
pii/tin	it	Codice di identificazione fiscale	MRTMTT25D09F205Z
pii/tin	li	Steuer-Identifikationsnummer	123456789012
pii/tin	lt	Mokesčių mokėtojo identifikacinis numeris	33309240064
pii/tin	lu	Steuer-Identifikationsnummer	1893120105732
pii/tin	lv	Nodokļu maksātāja identifikācijas numurs	121282-11212
pii/tin	mt	Numru ta' identifikazzjoni tal-persuna li tħallas it-taxxa	1234567A
pii/tin	nl	Fiscaal identificatienummer	174559434
pii/tin	no	Skatteidentifikasjonsnummer	912345678MVA
pii/tin	nz	IRD Number	123456785
pii/tin	pl	Numer identyfikacji podatnika	02070803628
pii/tin	pt	Número de identificação fiscal	299999998

pii/tin	ro	Număr de identificare fiscală	9000567890123
pii/tin	se	Skatteregistreringsnummer	870918-9990
pii/tin	si	Davčna številka davkoplačevalca	15012557
pii/tin	sk	DIČ	281203/054
pii/tin	tr	Vergi Kimlik Numarası	2460194630
pii/tin	us	ITIN	911-92-3333
pii/tin	za	Nomoro motšhelo	0001339050

The following table lists example landmarks and values for the CJKVT TIN grammar (`tin_cjkvt.ecr` or `tin_cjkvt.ejr`) for each supported country.

Type	Country	Landmark	Example
pii/tin	jp	法人番号	654327654322
pii/tin	tw	身份證字號	A123456789

Example Travel Numbers

The following table lists example landmarks and values for the travel grammar (`travel.ecr`).

Type	Country	Landmark	Example
pii/travel	us	passport card	C12345678

PII Grammar Customization

In cases where you find that the PII grammars miss particular matches in your input, you can customize them. This section describes the possible customizations.

The following grammars support customization:

- `address.ecr`
- `address_cjkvt.ecr`
- `name.ecr`
- `name_cjkvt.ecr`

The combined versions of these grammars also support customization. See Combined Grammars, on page 78.

> **NOTE:** It is technically possible to extend any public entity in a PII grammar, but it can involve a lot

of work. If you want to extend an entity that is not listed in the following list, see Modify Other Grammars and Entities, on page 79.

For each grammar that supports customization, you can customize the following entities:

- address
  - `pii/address/knowncity_headwords/CC`
  - `pii/address/knownstreet/CC`
- name
  - `pii/name/surname/nocontext/CC`
  - `pii/name/given_name/nocontext/CC`

In this list, CC means country code (for example: gb, us, nz). See Country and Language Support, on page 11.

You can use customizations to add entries that the existing entities do not match (such as unusual names). You might also use it if your data uses unusual separators and punctuation. The following sections provide examples of these changes.

TIP: When you customize an entity, you can either replace or extend the definition. For PII grammars, Micro Focus recommends that you only extend the entity definitions.

If you replace an entity, you are likely to miss matches or reduce performance. In addition, existing definitions cover many match cases that you might not consider, so there is a lot of value in using these definitions as a base.

Example 1: New Street Address

The following grammar definition below shows an example for extending `address.ecr`.

address_extended.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE grammars SYSTEM "../published/edk.dtd">
<grammars version="4.0">
   <include path="address.ecr"/>
   <grammar name="pii/address">

      <entity name="suffixes/gb" type="private">
         <entry headword="Cury"/>
         <entry headword="CURY"/>
      </entity>

      <entity name="knownstreet/gb" extend="append" type="private">
         <pattern>[A-Z][a-z]+ (?A:suffixes/gb)</pattern>
      </entity>
```

```
      <entity name="streetlocation/nocontext/gb" extend="append">
         <pattern score="0.75">(?A=STREET:(?A:knownstreet/gb))</pattern>
      </entity>

   </grammar>
</grammars>
```

This definition extends the `knownstreet/gb` and `streetlocation/nocontext/gb` entities in the PII address grammar:

- It adds *Cury* as a known street suffix.
- It extends the `knownstreet` entity to accept any two word street name that ends with the new *Cury* suffix.
- It extends the `streetlocation/nocontext/gb` entity to use the extended `knownstreet` entity, so that these changes take effect.

The result of these changes is that *Petty Cury* matches as a street location with a score of 0.75. Previously, it would not have matched at all.

> **TIP:** You do not need to redeclare the full address entity to use the extended `knownstreet` entity.
>
> For example, with these changes *123 Petty Cury, Cambridge CB4 0WZ* now matches `pii/address/gb` with a score of 1. Previously, this address would have matched, but with a lower score.
>
> When you add known street names or patterns for your country of interest, it improves scores for matches that contain these customizations.

Example 2: New Known City

The following grammar definition adds more known cities to `address.ecr`.

address_extended.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE grammars SYSTEM "../published/edk.dtd">
<grammars version="4.0">
   <include path="address.ecr"/>
   <grammar name="pii/address">

      <entity name="knowncity_headwords/gb" extend="append" type="private">
         <entry headword="Chesterton"/>
      </entity>

      <entity name="city/nocontext/gb" extend="append">
         <pattern>(?A=CITY:(?A^knowncity_headwords/gb))</pattern>
      </entity>

   </grammar>
</grammars>
```

This example definition:

- adds *Chesterton* to the `knowncity_headwords/gb entity`.
- extends the `city/nocontext/gb` entity to use the extended `knowncity` entity, so that the change takes effect.

The result of these changes is that *Chesterton* matches as a city with a score of 1. Previously, it would have matched as a speculative city name, with a lower score.

Again, you do not need to change the full address entity to pick up this new declaration. For example, *123 Main Street, Chesterton CB4 0WZ* now matches `pii/address/gb` with a score of 1, which is an improved score. Previously, it would have matched with a lower score, because the city was a speculative match.

> **TIP:** The definition for `city/nocontext/gb` uses the dynamic reference syntax when using the `knowncity_headwords/gb`; that is, `(?A^`. Micro Focus recommends this syntax for performance reasons when you refer to that entity, because the version of this entity for each country often contains several thousand entries.

To make both sets of changes for known streets and cities, merge the declarations in examples 1 and 2 into a single XML file.

Example 3: New Name and Custom Separator

Another way to use entity customizations is to declare patterns with custom separators. For example, if your input data contains unusual spacing or characters between entities, you can declare these in your entity extensions.

The following grammar definition extends `name.ecr`.

name_extended.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE grammars SYSTEM "../published/edk.dtd">
<grammars version="4.0">
   <include path="name.ecr"/>
   <grammar name="pii/name">

      <entity name="given_name/nocontext/gb" extend="append" case="insensitive">
         <entry headword="Fobo" score="2"/>
      </entity>

      <entity name="surname/nocontext/gb" extend="append" case="insensitive">
         <entry headword="Jobo" score="2"/>
      </entity>

      <entity name="gb" extend="append">
         <pattern>(?A=SURNAME:(?A:surname/nocontext/gb))@@(?A=FORENAME:(?A:given_
name/nocontext/gb))</pattern>
      </entity>
```

```
   </grammar>
</grammars>
```

This declaration makes two changes:

- It adds new entries for `given_name` and `surname`. This change allows *Fobo Jobo* to match as a name for the `gb` entity.
- It declares a new pattern for the gb entity, to match a name in reverse order, with the elements separated by a custom separator (two @ symbols). This change allows *Jobo@@Fobo* to match as a name.

> **TIP:** The grammar already handles hyphenated known names. For example, after this definition change, Eduction matches *Fobo-Fobo Jobo* with a score of 1, with no further changes required. You do not need to add hyphenated entries to the `given_name/nocontext` or `surname/nocontext` entities.

Combined Grammars

You can make the same extensions for the combined grammars. The following example updates the `combined_address` grammar to make the same changes as in Example 1: New Street Address, on page 75 and Example 2: New Known City, on page 76..

combined_address_extended.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE grammars SYSTEM "../published/edk.dtd">
<grammars version="4.0">
   <include path="combined_address.ecr"/>
   <grammar name="pii/address">

      <entity name="suffixes/gb" type="private">
         <entry headword="Cury"/>
         <entry headword="CURY"/>
      </entity>

      <entity name="knownstreet/gb" extend="append" type="private">
         <pattern>[A-Z][a-z]+ (?A:suffixes/gb)</pattern>
      </entity>

      <entity name="streetlocation/nocontext/all" extend="append">
         <pattern score="0.75">(?A=STREET:(?A:knownstreet/gb))</pattern>
      </entity>

      <entity name="knowncity_headwords/gb" extend="append" type="private">
         <entry headword="Chesterton"/>
      </entity>

      <entity name="city/nocontext/all" extend="append">
```

```
        <pattern>(?A=CITY:(?A^knowncity_headwords/gb))</pattern>
      </entity>

   </grammar>
</grammars>
```

NOTE: The public entities use `all` as the country code, while the private ones continue to use the appropriate country code.

Compile Custom Grammars

As with any Eduction grammar, Micro Focus recommends that you compile your grammar extensions before using them. You can use the edktool command-line tool to compile the XML file that contains your extension declarations into an ECR file.

For more information about compiling custom grammars, refer to the *Eduction User and Programming Guide*.

Modify Other Grammars and Entities

It is possible to extend any public entity in a PII grammar. However, you cannot use the various private entities that the public ones use in their definitions.

For entities in the simpler grammars such as driving or national ID, this might be less of a problem, as long as you know the format for the data portion of this entity. For example, you might want to add new landmarks to these entities, for example.

However, be aware that existing definitions account for factors such as varying spaces, and additional words between the landmark and the data. In this case, you must emulate this behavior in your extensions, which might take a lot of work.

In practice, Micro Focus recommends that you make a support request to make these changes to the official PII grammars, unless you need to add support in a very short time frame. The existing definitions provide a lot of value because they cover so many match cases, and you might miss these cases when you extend the public entities where these definitions are not available.

Validated ID Numbers

The script `pii_postprocessing.lua` (see Configure Post Processing, on page 16) includes steps to validate ID numbers that are found by Eduction. 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 (`health.ecr`)

pii/health/id/context/au	
pii/health/id/context/br	
pii/health/id/context/gb	
pii/health/id/context/fr	Validated using the INSEE checksum
pii/health/id/context/nz	

National ID numbers (`national_id.ecr` or `national_id_cjkvt.ecr`)		
pii/id/context/at	pii/id/nocontext/at	Only the SSN component is validated.
pii/id/context/be	pii/id/nocontext/be	
pii/id/context/bg	pii/id/nocontext/bg	
pii/id/context/ca	pii/id/nocontext/ca	
pii/id/context/ch	pii/id/nocontext/ch	
pii/id/context/cn	pii/id/nocontext/cn	
pii/id/context/cz	pii/id/nocontext/cz	
pii/id/context/ee	pii/id/nocontext/ee	
pii/id/context/es	pii/id/nocontext/es	
pii/id/context/fi	pii/id/nocontext/fi	
pii/id/context/fr	pii/id/nocontext/fr	
pii/id/context/gr	pii/id/nocontext/gr	Only the AMKA component is validated.
pii/id/context/hk	pii/id/nocontext/hk	
pii/id/context/hr	pii/id/nocontext/hr	
pii/id/context/hu	pii/id/nocontext/hu	Only the PIN component is validated.
pii/id/context/ie	pii/id/nocontext/ie	
pii/id/context/in	pii/id/nocontext/in	
pii/id/context/is	pii/id/nocontext/is	
pii/id/context/it	pii/id/nocontext/it	
pii/id/context/jp	pii/id/nocontext/jp	
pii/id/context/lt	pii/id/nocontext/lt	

pii/id/context/lu	pii/id/nocontext/lu	
pii/id/context/nl	pii/id/nocontext/nl	
pii/id/context/no	pii/id/nocontext/no	
pii/id/context/pl	pii/id/nocontext/pl	
pii/id/context/pt	pii/id/nocontext/pt	
pii/id/context/ro	pii/id/nocontext/ro	
pii/id/context/se	pii/id/nocontext/se	
pii/id/context/si	pii/id/nocontext/si	
pii/id/context/sk	pii/id/nocontext/sk	Only the Rodné číslo component is validated.
pii/id/context/th	pii/id/nocontext/th	
pii/id/context/tr	pii/id/nocontext/tr	
pii/id/context/tw	pii/id/nocontext/tw	
pii/id/context/za	pii/id/nocontext/za	

Tax ID numbers (`tin.ecr` or `tin_cjkvt.ecr`)		
pii/tin/context/at	pii/tin/nocontext/at	
pii/tin/context/au	pii/tin/nocontext/au	
pii/tin/context/be	pii/tin/nocontext/be	
pii/tin/context/bg	pii/tin/nocontext/bg	
pii/tin/context/br	pii/tin/nocontext/br	Cadastro de Pessoas Físicas (CPF) Cadastro Nacional de Pessoa Jurídica (CNPJ)
pii/tin/context/ca	pii/tin/nocontext/ca	
pii/tin/context/ch	pii/tin/nocontext/ch	Validation for AHV number, but not for UID-Nr.
pii/tin/context/cy	pii/tin/nocontext/cy	
pii/tin/context/cz	pii/tin/nocontext/cz	
pii/tin/context/de	pii/tin/nocontext/de	
pii/tin/context/dk	pii/tin/nocontext/dk	

pii/tin/context/ee	pii/tin/nocontext/ee	
pii/tin/context/es	pii/tin/nocontext/es	
pii/tin/context/fi	pii/tin/nocontext/fi	
pii/tin/context/fr	pii/tin/nocontext/fr	
pii/tin/context/hr	pii/tin/nocontext/hr	
pii/tin/context/hu	pii/tin/nocontext/hu	
pii/tin/context/ie	pii/tin/nocontext/ie	
pii/tin/context/is	pii/tin/nocontext/is	
pii/tin/context/it	pii/tin/nocontext/it	
pii/tin/context/jp	pii/tin/nocontext/jp	
pii/tin/context/lt	pii/tin/nocontext/lt	
pii/tin/context/lu	pii/tin/nocontext/lu	
pii/tin/context/mt	pii/tin/nocontext/mt	
pii/tin/context/nl	pii/tin/nocontext/nl	
pii/tin/context/nz	pii/tin/nocontext/nz	Inland Revenue Department Number
pii/tin/context/pl	pii/tin/nocontext/pl	
pii/tin/context/pt	pii/tin/nocontext/pt	
pii/tin/context/se	pii/tin/nocontext/se	
pii/tin/context/si	pii/tin/nocontext/si	
pii/tin/context/sk	pii/tin/nocontext/sk	
pii/tin/context/za	pii/tin/nocontext/za	

Machine readable passport numbers (`mrtd.ecr` and `mrtd_cjkvt.ecr`)
pii/mrtd/mrp
pii/mrtd/mrp_cjkvt
pii/mrtd/mrotd/td1
pii/mrtd/mrotd/td1_cjkvt

Driving License numbers (`driving_cjkvt.ecr`)	
pii/driving/context/tw	pii/driving/nocontext/tw

Ambiguous Entities

For some entities, IDOL PII Package cannot always unambiguously determine the country of origin for a value. For some of these cases, it can return an ambiguous result.

Cross-Language Passport Landmarks

The IDOL PII Package allows cross-language passport landmarks, so that it detects passport numbers provided in languages that do not belong to the associated passport country.

For example, the text "Oma passi on P 4366918" contains *passi*, which is Finnish for passport, and the number P 4366918, which is an Austrian passport number. The PII grammar identifies this as an Austrian passport number and returns the entity `pii/passport/at`.

In some cases, the country of origin is ambiguous. In this case, the grammar attempts to identify all possible countries and returns an entity with the label `ambiguous`.

Example 1

"Mon passeport est LA080402"

In this example, both the landmark text *passeport*, and the passport number could be from either Belgium, Canada, or Luxembourg. This example returns the entity `pii/passport/ambiguous/be_ca_lu` to represent all three possibilities.

Example 2

"Vegabref mitt er AA5275702"

In this example, *Vegabref* is Icelandic, but AA5275702 could be a passport number for several countries, not including Iceland. This example returns the entity `pii/passport/ambiguous/au_fi_ie_it_lv_pl_sk_si_gr_hu_ee_nl_de_us02` to represent all the possibilities.

> **NOTE:** The us02 option in this response means that this pattern scores 0.2 as a US passport pattern.

> **TIP:** In this example, if the text belonged to a language from one of the possible countries, the passport number would not be considered ambiguous. For example, "Oma passi on AA5275702" (where *passi* is Finnish for passport), returns the entity `pii/passport/fi`, because passi applies only to Finnish, and not to any of the other countries where the passport number is valid.

Ambiguous Driving License Matches

There are some countries that have some overlap in driving license number, and where the languages are the same it is not possible to identify which country a particular number comes from. In this case,

the grammar attempts to identify all possible countries and returns an entity with the label `ambiguous`. For example:

```
pii/driving/ambiguous/au_ie_us
pii/driving/ambiguous/au_nz_us
pii/driving/ambiguous/au_us
pii/driving/ambiguous/mt_us
```

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 PII Package 12.7)

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!