

Package ‘czso’

April 6, 2020

Title Use Open Data from the Czech Statistical Office in R

Version 0.2.3

Description Get programmatic access to the open data provided by the Czech Statistical Office (CZSO, <https://czso.cz>).

License MIT + file LICENSE

URL <https://github.com/petrbouchal/czso>

BugReports <https://github.com/petrbouchal/czso/issues>

Imports curl (*i*= 4.3),
dplyr (*i*= 0.8.3),
httr (*i*= 1.4.1),
jsonlite (*i*= 1.6),
lifecycle,
magrittr,
readr (*i*= 1.3.1),
rlang (*i*= 0.4.4),
stringi (*i*= 1.4.4),
stringr (*i*= 1.4.0),
tibble (*i*= 2.1.3),
tools (*i*= 3.6.0),
usethis (*i*= 1.5.1),
utils (*i*= 3.6.0)

RdMacros lifecycle

Encoding UTF-8

Language en

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.0.0

R topics documented:

<code>czso_get_catalogue</code>	2
<code>czso_get_dataset_doc</code>	3
<code>czso_get_dataset_metadata</code>	4
<code>czso_get_table</code>	5
<code>czso_get_table_schema</code>	6

get_catalogue	7
get_czso_catalogue	8
get_czso_dataset_doc	8
get_czso_dataset_metadata	9
get_czso_table_schema	9
get_table	10
Index	11

czso_get_catalogue	<i>Get catalogue of open CZSO datasets</i>
--------------------	--

Description

Retrieves a list of all CZSO's open datasets available from the Czech Open data catalogue.

Usage

```
czso_get_catalogue()
```

Details

Use the `dataset_id` column as an argument to `get_czso_table()`.

Value

a data frame with details on all CZSO datasets available in the Czech National Open Data Catalogue. The columns are fairly well described by their names, except:

- some columns contain IRIs instead of human readable text; still you can deduce the content from the IRI.
- the `spatial` column contains an IRI ending in the pattern `unit_type/unit_code`. The `unit_type` denotes what unit the data covers (scope/domain not granularity) and the second identifies the unit covered. The `unit_type` will usually be `"stat"` for "state" and the `unit_code` will be 1. The `unit_type` can also be `"KR"` for region or `"OB"` for municipality, or `"OK"` for district. In that case, the `unit_code` will be a code of that unit.
- `page` points to the documentation, i.e. methodology notes for the dataset.

See Also

Other Core workflow: [czso_get_table\(\)](#), [get_table\(\)](#)

Examples

```
czso_get_catalogue()
```

czso_get_dataset_doc *Get documentation for CZSO dataset*

Description

Retrieves the URL/downloads the file containing the documentation of the dataset, in the required format.

Usage

```
czso_get_dataset_doc(  
  dataset_id,  
  action = c("return", "open", "download"),  
  destfile = NULL,  
  format = c("html", "pdf", "word")  
)
```

Arguments

dataset_id	Dataset ID
action	Whether to return URL (the default), download the file, or open the URL in the default web browser.
destfile	Where to save the file. Only used if if action = download.
format	What file format to access: html (the default), pdf, or word.

Details

The document to which this functions provides access contains methodological background on the specified dataset and is identified by the `schema` field in the list returned by `czso_get_dataset_metadata()`

Value

if `action = download`, the path to the downloaded file; file URL otherwise.

See Also

Other Additional tools: [czso_get_dataset_metadata\(\)](#), [czso_get_table_schema\(\)](#), [get_czso_dataset_doc\(\)](#), [get_czso_dataset_metadata\(\)](#), [get_czso_table_schema\(\)](#)

Examples

```
czso_get_dataset_doc("110080")
```

`czso_get_dataset_metadata`*Get dataset metadata*

Description

Get metadata from CZSO API, which can be somewhat more detailed/readable than what is provided in the dataset's entry in the output of `czso_get_dataset()`.

Usage

```
czso_get_dataset_metadata(dataset_id)
```

Arguments

`dataset_id` Dataset ID

Details

As far as I can tell there is no way to get the metadata in English, though some key datasets, such as codelists, do have English-language documentation. See `czso_get_table()` for how to access English-language codelists (registers).

Value

a list with elements named in English, where the names are mostly self-explanatory. So are the contents where these are dates; title, description, notes and tags only exist in Czech as far as I know. Some fields merit explanation:

- `resources`: a list of files available to download in this dataset
- `frequency`: see https://project-open-data.cio.gov/iso8601_guidance/ for a key
- `ruian_type`: what type of spatial unit the data covers (spatial domain/extent/scope, not granularity). ST means "state" (this is almost always the case), "KR" means region (kraj), "OK" district (okres), "OB" municipality (obec); "RS" cohesion region (region soudržnosti, larger than region)
- `ruian_code`: the code of the unit the data covers as per the RUIAN taxonomy
- `schema` points to documentation while `describedBy` points to the technical schema in JSON or XML.

See Also

Other Additional tools: [czso_get_dataset_doc\(\)](#), [czso_get_table_schema\(\)](#), [get_czso_dataset_doc\(\)](#), [get_czso_dataset_metadata\(\)](#), [get_czso_table_schema\(\)](#)

Examples

```
czso_get_dataset_metadata("110080")
```

czso_get_table	<i>Retrieve and read dataset from CZSO</i>
----------------	--

Description

Downloads and reads dataset identified by `dataset_id`. Unzips if necessary, but only loads CSV files, otherwise returns the path to the downloaded file. Converts types of columns where known, e.g. value columns to numeric.

Usage

```
czso_get_table(dataset_id, force_redownload = FALSE, resource_num = 1)
```

Arguments

<code>dataset_id</code>	a character. Found in the <code>czso_id</code> column of data frame returned by <code>get_catalogue()</code> .
<code>force_redownload</code>	integer. Whether to redownload data source file even if already cached. Defaults to <code>FALSE</code> .
<code>resource_num</code>	integer. Order of resource in resource list for the given dataset. Defaults to 1, the normal value for CZSO datasets.

Details

Structure of the output tibble:

CZSO provides its open data as tidy data, so each row only contains one value in the `hodnota` column and the remaining columns give details on how that value is defined. See "Included columns" below on how these work.

Data types:

The schema of the dataset is not yet used, so some columns may be mistyped and are by default returned as character vectors.

Included columns:

The range of columns present in the output vary from one dataset to another, so the package does not attempt to provide English-language names for the known subset, as that would result in a jumble of Czenlish.

Instead, here is a guide to some of the common column names you will encounter:

- `idhod`: a unique ID of the value in the CZSO databse. This does not allow you to link to any other (meta)data as far as I know, but it does provide unique identification should you need it.
- `hodnota`: the value.
- `stapro_kod`: code of the statistic/indicator/variable as listed. in the SMS UKAZ register (<https://www.czso.cz/csu/czso/statistical-variables-indicators>); this one has Czech-English documentation - access this by clicking the UK flag top right.
- `rok` denotes year as YYYY.
- `ctvrtleti` denotes quarter if available.

Other metadata will come in the form {variable}_{txt—cis—kod}. The `.txt` column holds the Czech text name for the category. The `.cis` column holds the ID of the codelist (register) you need to decode the code in `.kod`. The English codelists are at <http://apl.czso.cz/iSMS/en/cislist.jsp>, Czech ones at <http://apl.czso.cz/iSMS/cs/cislist.jsp>. You can find the Czech-language codelists in the catalogue retrieved with `czso_get_catalogue()`; the English ones can also be retrieved from the link above using a permalink URL.

Units are denoted in a separate column.

A helper on common breakdowns with their associated columns:

- `uzemi`: territory
- `vek`: age
- `pohlavi`: gender

NAs in "breakdown" columns (e.g. gender or age) denote the total.

Value

a tibble, or vector of file paths if file is not CSV or if there are multiple files in the dataset. See Details on the columns contained in the tibble

Note

Do not use this for harvesting datasets from CZSO en masse.

See Also

Other Core workflow: [czso_get_catalogue\(\)](#), [get_table\(\)](#)

Examples

```
czso_get_table("110080")
```

`czso_get_table_schema` *Get CZSO table schema*

Description

Retrieves and parses the schema for the table identified by `dataset_id` and `resource_num`.

Usage

```
czso_get_table_schema(dataset_id, resource_num = 1)
```

Arguments

`dataset_id` Dataset ID
`resource_num` Resource number, typically 1 in CZSO (the default)

Details

Currently only handles JSON schema files for CSV files. If the schema is a different format, an error is returned pointing the user to the URL of the file.

Value

a tibble with a description of the table columns, with the following items:

- name: the column name.
- titles: usually the duplicate of name
- dc:description: a Czech-language description of the column
- required: whether the column is required
- datatype: the data type of the column; either "number" or "string"

See Also

Other Additional tools: [czso_get_dataset_doc\(\)](#), [czso_get_dataset_metadata\(\)](#), [get_czso_dataset_doc\(\)](#), [get_czso_dataset_metadata\(\)](#), [get_czso_table_schema\(\)](#)

Examples

```
czso_get_table_schema("110080")
```

get_catalogue

Deprecated: use czso_get_catalogue() instead

Description

Soft-deprecated

Usage

```
get_catalogue()
```

Value

a tibble

Examples

```
# see `czso_get_catalogue()`
```

get_czso_catalogue *Deprecated, use czso_get_catalogue() instead.*

Description

Soft-deprecated

Usage

```
get_czso_catalogue()
```

Value

a tibble

Examples

```
# see `czso_get_catalogue()`
```

get_czso_dataset_doc *Deprecated: use czso_get_dataset_doc() instead*

Description

Soft-deprecated

Usage

```
get_czso_dataset_doc(
  dataset_id,
  action = c("return", "open", "download"),
  destfile = NULL,
  format = c("html", "pdf", "word")
)
```

Arguments

dataset_id	Dataset ID
action	Whether to return URL (the default), download the file, or open the URL in the default web browser.
destfile	Where to save the file. Only used if if action = download.
format	What file format to access: html (the default), pdf, or word.

Value

a list

See Also

Other Additional tools: [czso_get_dataset_doc\(\)](#), [czso_get_dataset_metadata\(\)](#), [czso_get_table_schema\(\)](#), [get_czso_dataset_metadata\(\)](#), [get_czso_table_schema\(\)](#)

`get_czso_dataset_metadata`*Deprecated, use `czso_get_catalogue()` instead.*

Description

Soft-deprecated

Usage

```
get_czso_dataset_metadata(dataset_id)
```

Arguments

`dataset_id` Dataset ID

Value

a list

See Also

Other Additional tools: [czso_get_dataset_doc\(\)](#), [czso_get_dataset_metadata\(\)](#), [czso_get_table_schema\(\)](#), [get_czso_dataset_doc\(\)](#), [get_czso_table_schema\(\)](#)

`get_czso_table_schema` *Deprecated: use `czso_get_table_schema()` instead*

Description

Soft-deprecated

Usage

```
get_czso_table_schema(dataset_id, resource_num)
```

Arguments

`dataset_id` Dataset ID
`resource_num` Resource number, typically 1 in CZSO (the default)

Value

a list

See Also

Other Additional tools: [czso_get_dataset_doc\(\)](#), [czso_get_dataset_metadata\(\)](#), [czso_get_table_schema\(\)](#), [get_czso_dataset_doc\(\)](#), [get_czso_dataset_metadata\(\)](#)

get_table *Deprecated: use czso_get_table() instead.*

Description

Soft-deprecated.

Usage

```
get_table(dataset_id, resource_num = 1, force_redownload = FALSE)
```

Arguments

dataset_id a character. Found in the `czso_id` column of data frame returned by `get_catalogue()`.

resource_num integer. Order of resource in resource list for the given dataset. Defaults to 1, the normal value for CZSO datasets.

force_redownload integer. Whether to redownload data source file even if already cached. Defaults to FALSE.

Value

a tibble

See Also

Other Core workflow: [czso_get_catalogue\(\)](#), [czso_get_table\(\)](#)

Examples

```
# see `czso_get_table()`
```

Index

`czso_get_catalogue`, [2](#), [6](#), [10](#)
`czso_get_dataset_doc`, [3](#), [4](#), [7-9](#)
`czso_get_dataset_metadata`, [3](#), [4](#), [7-9](#)
`czso_get_table`, [2](#), [5](#), [10](#)
`czso_get_table_schema`, [3](#), [4](#), [6](#), [8](#), [9](#)

`get_catalogue`, [7](#)
`get_czso_catalogue`, [8](#)
`get_czso_dataset_doc`, [3](#), [4](#), [7](#), [8](#), [9](#)
`get_czso_dataset_metadata`, [3](#), [4](#), [7-9](#), [9](#)
`get_czso_table_schema`, [3](#), [4](#), [7-9](#), [9](#)
`get_table`, [2](#), [6](#), [10](#)