

Package ‘statnipokladna’

August 21, 2024

Title Use Data from the Czech Public Finance Database

Version 0.7.4

Description Get programmatic access to data from the Czech public budgeting and accounting database, Státní pokladna
<<https://monitor.statnipokladna.cz/>>.

License MIT + file LICENSE

URL <https://github.com/petrbouchal/statnipokladna>,
<https://petrbouchal.xyz/statnipokladna/>

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

Depends R (>= 3.5.0)

Imports cli, curl (>= 4.3), dplyr (>= 1.0.0), httr (>= 1.4.1),
jsonlite, lifecycle (>= 1.0.1), lubridate (>= 1.7.4), magrittr,
purrr (>= 0.3.2), readr (>= 1.3.1), rlang (>= 0.4.0), stringi
(>= 1.4.3), stringr (>= 1.4.0), tibble (>= 2.1.3), tidyr (>=
1.0.0), utils (>= 3.6.0), xml2 (>= 1.2.2)

Suggests knitr (>= 1.30), markdown (>= 1.1), ragg (>= 0.4.0),
rmarkdown, testthat (>= 2.1.0), tidyselect (>= 1.2.0)

VignetteBuilder knitr

Encoding UTF-8

Language en

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Author Petr Bouchal [aut, cre] (<<https://orcid.org/0000-0002-0471-716X>>)

Maintainer Petr Bouchal <pbouchal@gmail.com>

Repository CRAN

Date/Publication 2024-08-21 11:00:02 UTC

Contents

add_codelist	2
get_codelist	3
sp_add_codelist	4
sp_codelists	5
sp_datasets	6
sp_get_codelist	7
sp_get_codelist_file	8
sp_get_codelist_url	9
sp_get_dataset	9
sp_get_dataset_doc	11
sp_get_dataset_url	11
sp_get_table_file	12
sp_load_codelist	13
sp_load_table	14
sp_tables	14
Index	16

add_codelist	<i>Deprecated: Add codelist data to downloaded data</i>
--------------	---

Description

Deprecated, use `sp_add_codelist()` instead.

Usage

```
add_codelist(
  data,
  codelist = NULL,
  period_column = .data$vykaz_date,
  redownload = FALSE,
  dest_dir = NULL
)
```

Arguments

<code>data</code>	a data frame returned by <code>sp_get_table()</code> .
<code>codelist</code>	The codelist to add. Either a character vector of length one (see <code>sp_tables</code> for possible values), or a data frame returned by <code>sp_get_codelist()</code> .
<code>period_column</code>	Unquoted column name of column identifying the data period in <code>data</code> . Leave to default if you have not changed the data object returned by <code>sp_get_table()</code> .
<code>redownload</code>	Redownload even if file has already been downloaded? Defaults to <code>FALSE</code> .

`dest_dir` character. Directory in which downloaded files will be stored. If left unset, will use the `statnipokladna.dest_dir` option if the option is set, and `tempdir()` otherwise. Will be created if it does not exist.

Details

[Deprecated]

Value

A [tibble](#) of same length as `data`, with added columns from `codelist`. See [Details](#).

See Also

Other Core workflow: [get_codelist\(\)](#), [sp_add_codelist\(\)](#), [sp_get_codelist\(\)](#), [sp_get_dataset\(\)](#), [sp_get_table\(\)](#)

<code>get_codelist</code>	<i>Deprecated: Get codelist</i>
---------------------------	---------------------------------

Description

Deprecated: use `sp_get_codelist()`

Usage

```
get_codelist(codelist_id, n = NULL, dest_dir = NULL, redownload = FALSE)
```

Arguments

<code>codelist_id</code>	A codelist ID. See <code>id</code> column in <code>sp_codelists</code> for a list of available codelists.
<code>n</code>	Number of rows to return. Default (<code>NULL</code>) means all. Useful for quickly inspecting a codelist.
<code>dest_dir</code>	character. Directory in which downloaded files will be stored. If left unset, will use the <code>statnipokladna.dest_dir</code> option if the option is set, and <code>tempdir()</code> otherwise. Will be created if it does not exist.
<code>redownload</code>	Redownload even if file has already been downloaded? Defaults to <code>FALSE</code> .

Details

[Deprecated]

Value

A [tibble](#)

See Also

Other Core workflow: [add_codelist\(\)](#), [sp_add_codelist\(\)](#), [sp_get_codelist\(\)](#), [sp_get_dataset\(\)](#), [sp_get_table\(\)](#)

sp_add_codelist	<i>Add codelist data to downloaded data</i>
-----------------	---

Description

Joins a provided codelist, or downloads and processes one if necessary, and adds it to the data.

Usage

```
sp_add_codelist(
  data,
  codelist = NULL,
  period_column = .data$vykaz_date,
  by = NULL,
  redownload = FALSE,
  dest_dir = NULL
)
```

Arguments

data	a data frame returned by <code>sp_get_table()</code> .
codelist	The codelist to add. Either a character vector of length one (see <code>sp_tables</code> for possible values), or a data frame returned by <code>sp_get_codelist()</code> .
period_column	Unquoted column name of column identifying the data period in data. Leave to default if you have not changed the data object returned by <code>sp_get_table()</code> .
by	character. Columns by which to join the codelist. Same form as for <code>dplyr::left_join()</code> ...
redownload	Redownload even if file has already been downloaded? Defaults to FALSE.
dest_dir	character. Directory in which downloaded files will be stored. If left unset, will use the <code>statnipokladna.dest_dir</code> option if the option is set, and <code>tempdir()</code> otherwise. Will be created if it does not exist.

Details

The data argument should be a data frame produced by `sp_get_table()` If this is true, the `period_column` argument is not needed. The `codelist` argument, if a data frame, should be a data frame produced by `sp_get_codelist()`. Specifically, it assumes it contains the following columns:

- `start_date`, a date
- `end_date`, a date
- column with the code, character usually named the same as the codelist

#' You can usually tell which codelist you need from the name of the column whose code you are looking to expand, e.g. the codes in column `paragraf` can be expanded by codelist `paragraf`.

The function filters the codelist to obtain a set of entries relevant to the time period of data. If data contains tables for multiple periods, this is handled appropriately. Codelist-originating columns in the resulting data frame are renamed so they do not interfere with joining additional codelists, perhaps in a single pipe call.

Note that some codelists are "secondary" and can only be joined onto other codelists. If a codelist does not join using `sp_add_codelis()`, store the output of `sp_get_codelist()` and join it manually using `dplyr`.

Value

A [tibble](#) of same length as `data`, with added columns from `codelist`. See [Details](#).

See Also

Other Core workflow: [add_codelist\(\)](#), [get_codelist\(\)](#), [sp_get_codelist\(\)](#), [sp_get_dataset\(\)](#), [sp_get_table\(\)](#)

Examples

```
## Not run:
sp_get_table("budget-central", 2017) %>%
  sp_add_codelist("polozka") %>%
  sp_add_codelist("paragraf")

pol <- sp_get_codelist("paragraf")
par <- sp_get_codelist("polozka")

sp_get_table("budget-central", 2017) %>%
  sp_add_codelist(pol) %>%
  sp_add_codelist(par)

## End(Not run)
```

sp_codelists

List of available codelists

Description

Contains IDs and names of all (most) available codelists that can be retrieved by `sp_get_codelist`.

Usage

```
sp_codelists
```

Format

A data frame with 27 rows and 2 variables:

id character. ID, used as codelist_id argument in sp_get_codelist.

name character. Short name, mostly corresponds to title used on statnipokladna.cz.

Details

The id is to be used as the codelist_id parameter in sp_get_codelist. See <https://monitor.statnipokladna.cz/datovy-katalog/ciselnyky> for a more detailed descriptions and a GUI for exploring the lists.

See Also

Other Lists of available entities: [sp_datasets](#), [sp_tables](#)

sp_datasets

List of available datasets

Description

Contains IDs and names of all available datasets that can be retrieved by get_dataset.

Usage

```
sp_datasets
```

Format

A data frame with 9 rows and 3 variables:

id character. Dataset ID, used as dataset_id argument to sp_get_dataset.

name character. Dataset name, mostly corresponds to title on the statnipokladna GUI.

Details

See <https://monitor.statnipokladna.cz/datovy-katalog/transakcni-data> for a more detailed descriptions of the datasets.

See Also

Other Lists of available entities: [sp_codelists](#), [sp_tables](#)

sp_get_codelist	<i>Get codelist</i>
-----------------	---------------------

Description

Downloads and processes codelist identified by `codelist_id`. See `sp_codelists` for a list of available codelists with their IDs and names.

Usage

```
sp_get_codelist(codelist_id, n = NULL, dest_dir = NULL, redownload = FALSE)
```

Arguments

<code>codelist_id</code>	A codelist ID. See <code>id</code> column in <code>sp_codelists</code> for a list of available codelists.
<code>n</code>	Number of rows to return. Default (NULL) means all. Useful for quickly inspecting a codelist.
<code>dest_dir</code>	character. Directory in which downloaded files will be stored. If left unset, will use the <code>statnipokladna.dest_dir</code> option if the option is set, and <code>tempdir()</code> otherwise. Will be created if it does not exist.
<code>redownload</code>	Redownload even if file has already been downloaded? Defaults to FALSE.

Details

You can usually tell which codelist you need from the name of the column whose code you are looking to expand, e.g. the codes in column `paragraf` can be expanded by codelist `paragraf`.

The processing ensures that the resulting codelist can be correctly joined to the data, automatically using `sp_add_codelist()` or manually. The entire codelist is downloaded and not filtered for any particular date.

Codelist XML files are stored in a temporary directory as determined by `tempdir()` and persist per session to avoid redownloads.

Value

a [tibble](#)

See Also

Other Core workflow: [add_codelist\(\)](#), [get_codelist\(\)](#), [sp_add_codelist\(\)](#), [sp_get_dataset\(\)](#), [sp_get_table\(\)](#)

Examples

```
## Not run:  
sp_get_codelist("paragraf")  
  
## End(Not run)
```

sp_get_codelist_file *Download a codelist XML file*

Description

This is normally called inside `sp_get_codelist()` but can be used separately if finer-grained control of intermediate outputs is needed, e.g. in a `{targets}` workflow.

Usage

```
sp_get_codelist_file(
  codelist_id = NULL,
  url = NULL,
  dest_dir = NULL,
  redownload = FALSE
)
```

Arguments

<code>codelist_id</code>	A codelist ID. See <code>id</code> column in <code>sp_codelists</code> for a list of available codelists.
<code>url</code>	DESCRIPTION. Either this or <code>codelist_id</code> must be set. If both are set, <code>url</code> wins.
<code>dest_dir</code>	character. Directory in which downloaded files will be stored. If left unset, will use the <code>statnipokladna.dest_dir</code> option if the option is set, and <code>tempdir()</code> otherwise. Will be created if it does not exist.
<code>redownload</code>	Redownload even if file has already been downloaded? Defaults to <code>FALSE</code> .

Value

path to XML file; character vector of length one.

See Also

Other Detailed workflow: [sp_get_codelist_url\(\)](#), [sp_get_dataset_url\(\)](#), [sp_get_table_file\(\)](#), [sp_load_codelist\(\)](#), [sp_load_table\(\)](#)

Examples

```
## Not run:
sp_get_codelist_file("druhuj")
codelist_url <- sp_get_codelist_url("druhuj")
sp_get_codelist_file(url = codelist_url)

## End(Not run)
```

sp_get_codelist_url *Get URL of a given codelist*

Description

This is normally called inside `sp_get_codelist()` but can be used separately if finer-grained control of intermediate outputs is needed, e.g. in a `{targets}` workflow.

Usage

```
sp_get_codelist_url(codelist_id, check_if_exists = TRUE)
```

Arguments

codelist_id	DESCRIPTION.
check_if_exists	Whether to check that the URL works (HTTP 200).

Value

character vector of length one containing URL

See Also

Other Detailed workflow: [sp_get_codelist_file\(\)](#), [sp_get_dataset_url\(\)](#), [sp_get_table_file\(\)](#), [sp_load_codelist\(\)](#), [sp_load_table\(\)](#)

Examples

```
## Not run:  
sp_get_codelist_url("ucjed", FALSE)  
if(FALSE) sp_get_codelist_url("ucjed_wrong", TRUE) # fails, invalid codelist  
  
## End(Not run)
```

sp_get_dataset *Retrieve dataset from statnipokladna*

Description

Downloads ZIP archives for a given dataset. If year or month have length > 1, gets all combinations.

Usage

```
sp_get_dataset(
  dataset_id,
  year,
  month = 12,
  dest_dir = NULL,
  redownload = FALSE
)
```

Arguments

dataset_id	A dataset ID. See id column in sp_datasets for a list of available datasets
year	year, numeric vector of length <= 1 (can take multiple values), 2015-2019 for some datasets, 2010-2020 for others. Defaults to 2018. (see Details for how to work with data across time periods.)
month	month, numeric vector of length <= 1 (can take multiple values). Must be between 1 and 12. Defaults to 12. (see Details for how to work with data across time periods.)
dest_dir	character. Directory in which downloaded files will be stored. If left unset, will use the statnipokladna.dest_dir option if the option is set, and tempdir() otherwise. Will be created if it does not exist.
redownload	Redownload even if file has already been downloaded? Defaults to FALSE.

Details

Files are stored in a temp folder as determined by tempdir() or the dest_dir param or the statnipokladna.dest_dir option. and further sorted into subdirectories by dataset, year and month. If saved to tempdir() (the default), downloaded files per session to avoid redownloads.

How data for different time periods is exported differs by dataset. This has significant implications for how you get to usable full-year numbers or time series in different tables. See vignette("statnipokladna") for details on this.

Value

character string with complete paths to downloaded ZIP archives.

See Also

Other Core workflow: [add_codelist\(\)](#), [get_codelist\(\)](#), [sp_add_codelist\(\)](#), [sp_get_codelist\(\)](#), [sp_get_table\(\)](#)

Examples

```
## Not run:
budget_2018 <- sp_get_dataset("finm", 2018)
budget_mid2018 <- sp_get_dataset("finm", 2018, 6)

## End(Not run)
```

sp_get_dataset_doc *Get dataset documentation*

Description

Downloads XLS file with dataset documentation, or opens link to this file in browser.

Usage

```
sp_get_dataset_doc(dataset_id, dest_dir = NULL, download = TRUE)
```

Arguments

dataset_id	dataset ID. See sp_datasets.
dest_dir	character. Directory in which downloaded files will be stored. If left unset, will use the statnipokladna.dest_dir option if the option is set, and tempdir() otherwise. Will be created if it does not exist.
download	Whether to download (the default) or open link in browser.

Value

(invisible) path to file if download = TRUE, URL otherwise

See Also

Other Utilities: [sp_get_codelist_viewer\(\)](#)

Examples

```
## Not run:  
sp_get_dataset_doc("finm")  
  
## End(Not run)
```

sp_get_dataset_url *Get URL of dataset*

Description

Useful for workflows where you want to keep track of URLs and intermediate files, rather than having all steps performed by one function.

Usage

```
sp_get_dataset_url(dataset_id, year, month = 12, check_if_exists = TRUE)
```

Arguments

dataset_id	Dataset ID. See id column in sp_datasets for a list of available codelists.
year	year, numeric vector of length ≤ 1 (can take multiple values), 2015-2019 for some datasets, 2010-2020 for others. (see Details for how to work with data across time periods.)
month	month, numeric vector of length ≤ 1 (can take multiple values). Must be between 1 and 12. Defaults to 12. (see Details for how to work with data across time periods.)
check_if_exists	Whether to check that the URL works (HTTP 200).

Value

a character vector of length one, containing a URL

See Also

Other Detailed workflow: [sp_get_codelist_file\(\)](#), [sp_get_codelist_url\(\)](#), [sp_get_table_file\(\)](#), [sp_load_codelist\(\)](#), [sp_load_table\(\)](#)

Examples

```
## Not run:
sp_get_dataset_url("finm", 2018, 6, FALSE)
sp_get_dataset_url("finm", 2029, 6, FALSE) # works but returns invalid URL
if(FALSE) sp_get_dataset_url("finm_wrong", 2018, 6, TRUE) # fails, invalid dataset ID
if(FALSE) sp_get_dataset_url("finm", 2022, 6, TRUE) # fails, invalid time period

## End(Not run)
```

sp_get_table_file *Get path to a CSV file containing a table.*

Description

This is normally called inside sp_get_table() but can be used separately if finer-grained control of intermediate outputs is needed, e.g. in a {targets} workflow.

Usage

```
sp_get_table_file(table_id, dataset_path, reunzip = FALSE)
```

Arguments

table_id	Table ID; see id column in sp_tables for a list of available codelists.
dataset_path	Path to downloaded dataset, as output by sp_get_dataset()
reunzip	Whether to overwrite existing CSV files by unzipping the archive downloaded by sp_get_dataset(). Defaults to FALSE.

Value

Character vector of length one - a path.

See Also

Other Detailed workflow: [sp_get_codelist_file\(\)](#), [sp_get_codelist_url\(\)](#), [sp_get_dataset_url\(\)](#), [sp_load_codelist\(\)](#), [sp_load_table\(\)](#)

Examples

```
## Not run:
ds <- sp_get_dataset("rozv", 2018, 12)
sp_get_table_file("balance-sheet", ds)

## End(Not run)
```

sp_load_codelist	<i>Load codelist into a tibble from XML file</i>
------------------	--

Description

This is normally called inside `sp_get_codelist()` but can be used separately if finer-grained control of intermediate outputs is needed, e.g. in a `{targets}` workflow.

Usage

```
sp_load_codelist(path, n = NULL)
```

Arguments

path	Path to a file as returned by <code>sp_get_codelist_file()</code>
n	Number of rows to return. Default (NULL) means all. Useful for quickly inspecting a codelist.

Value

a [tibble](#)

See Also

Other Detailed workflow: [sp_get_codelist_file\(\)](#), [sp_get_codelist_url\(\)](#), [sp_get_dataset_url\(\)](#), [sp_get_table_file\(\)](#), [sp_load_table\(\)](#)

Examples

```
## Not run:
cf <- sp_get_codelist_file("druhuj")
sp_load_codelist(cf)

## End(Not run)
```

sp_load_table	<i>Load a statnipokladna table from a CSV file</i>
---------------	--

Description

This is normally called inside `sp_get_table()` but can be used separately if finer-grained control of intermediate outputs is needed, e.g. in a `{targets}` workflow.

Usage

```
sp_load_table(path, ico = NULL)
```

Arguments

path	path to a CSV file, as output by <code>sp_get_table_file()</code> .
ico	Organisation ID to filter by, if supplied.

Value

a [tibble](#). See help for `sp_get_table()` for a key to the columns.

See Also

Other Detailed workflow: [sp_get_codelist_file\(\)](#), [sp_get_codelist_url\(\)](#), [sp_get_dataset_url\(\)](#), [sp_get_table_file\(\)](#), [sp_load_codelist\(\)](#)

Examples

```
## Not run:
ds <- sp_get_dataset("rozv", 2018, 12)
tf <- sp_get_table_file("balance-sheet", ds)
sp_load_table(tf)

## End(Not run)
```

sp_tables	<i>List of available tables (PARTIAL)</i>
-----------	---

Description

Contains IDs and names of all available tables that can be retrieved by `sp_get_table`. Look inside the XLS documentation for each dataset at <https://monitor.statnipokladna.cz/datovy-katalog/transakcni-data> to see more detailed descriptions. Note that tables do not correspond to the `tabulka/vtab` attribute of the tables, they represent files inside datasets.

Usage

`sp_tables`

Format

A data frame with 2 rows and 4 variables:

`id` character Table id, used as `table_id` argument to `sp_get_table`.

`dataset_id` integer Table number.

`czech_name` character Czech name of the table.

`note` character Note.

See Also

Other Lists of available entities: [sp_codelists](#), [sp_datasets](#)

Index

* Core workflow

- add_codelist, 2
- get_codelist, 3
- sp_add_codelist, 4
- sp_get_codelist, 7
- sp_get_dataset, 9

* Detailed workflow

- sp_get_codelist_file, 8
- sp_get_codelist_url, 9
- sp_get_dataset_url, 11
- sp_get_table_file, 12
- sp_load_codelist, 13
- sp_load_table, 14

* Lists of available entities

- sp_codelists, 5
- sp_datasets, 6
- sp_tables, 14

* Utilities

- sp_get_dataset_doc, 11

* datasets

- sp_codelists, 5
- sp_datasets, 6
- sp_tables, 14

add_codelist, 2, 4, 5, 7, 10

get_codelist, 3, 3, 5, 7, 10

sp_add_codelist, 3, 4, 4, 7, 10

sp_codelists, 5, 6, 15

sp_datasets, 6, 6, 15

sp_get_codelist, 3–5, 7, 10

sp_get_codelist_file, 8, 9, 12–14

sp_get_codelist_url, 8, 9, 12–14

sp_get_codelist_viewer, 11

sp_get_dataset, 3–5, 7, 9

sp_get_dataset_doc, 11

sp_get_dataset_url, 8, 9, 11, 13, 14

sp_get_table, 3–5, 7, 10

sp_get_table_file, 8, 9, 12, 12, 13, 14

sp_load_codelist, 8, 9, 12, 13, 13, 14

sp_load_table, 8, 9, 12, 13, 14

sp_tables, 6, 14

tibble, 3, 5, 7, 13, 14