Package 'statnipokladna'

April 11, 2020

Title Use Data from the Czech Public Finance Database

Version 0.5.7

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

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

Depends R (>= 3.5.0)

Imports curl (>= 4.3),
dplyr (>= 0.8.3),
httr (>= 1.4.1),
jsonlite,
lifecycle,
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),
usethis (>= 1.5.1),
utils (>= 3.6.0),
xml2 (>= 1.2.2)

Suggests knitr,
rmarkdown,
testthat (>= 2.1.0)

VignetteBuilder knitr

RdMacros lifecycle

Encoding UTF-8

Language en

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.0.2

R topics documented:

---

add_codelist   *Deprecated: Add codelist data to downloaded data*

Description

Deprecated, use sp_add_codelist() instead.

Usage

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

Arguments

data	a data frame returned by sp_get_table().
codelist	The codelist to add. Either a character vector of length one (see sp_tables for possible values), or a data frame returned by sp_get_codelist().
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 sp_get_table().
redownload	Redownload even if file has already been downloaded? Defaults to FALSE.
dest_dir	character. Directory in which downloaded files will be stored. Defaults to tempdir().

Details

Soft-deprecated

Value

A data frame 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()

---

get_codelist   *Deprecated: Get codelist*

Description

Deprecated: use sp_get_codelist()

Usage

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

Arguments

codelist_id   A codelist ID. See id column in sp_codelists for a list of available codelists.

n   Number of rows to return. Default (NULL) means all. Useful for quickly in-
                specting a codelist.

dest_dir   character. Directory in which downloaded files will be stored. Defaults to
                tempdir().

redownload   Redownload even if file has already been downloaded? Defaults to FALSE.

Value

A tibble

See Also

Other Core workflow: add_codelist(), sp_add_codelist(), sp_get_codelist(), sp_get_dataset(),
sp_get_table()

get_dataset *Deprecated: Retrieve and read dataset from statnipokladna*

Description

Deprecated, use sp_get_dataset() instead.

Usage

```
get_dataset(
  dataset_id,
  year = 2019,
  month = 12,
  dest_dir = tempdir(),
  redownload = FALSE
)
```

Arguments

dataset_id	A dataset ID. See id column in sp_datasets for a list of available codelists.
year	year, numeric, 2015-2018 for some datasets, 2010-2019 for others. Defaults to 2018. (see Details for how to work with data across time periods.)
month	month, numeric. 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. Defaults to tempdir(). Will be created if it does not exist.
redownload	Redownload even if file has already been downloaded? Defaults to FALSE.

Details

Soft-deprecated

Value

character (link) if download = TRUE, nothing otherwise.

See Also

Other Utilities: [sp_get_codelist_viewer](), [sp_get_dataset_doc]()

---

get_dataset_doc   *Deprecated: Get dataset documentation*

Description

Deprecated, use `sp_get_dataset_doc()` instead.

Usage

```
get_dataset_doc(dataset_id, dest_dir = ".", download = TRUE)
```

Arguments

dataset_id	dataset ID. See `sp_datasets`.
dest_dir	character. Where the file should be written. Defaults to working directory.
download	Whether to download (the default) or open link in browser.

Details

Soft-deprecated

Value

a tibble

---

get_table   *Deprecated: Get a statnipokladna table*

Description

Deprecated, use `sp_get_table()` instead.

Usage

```
get_table(
  table_id,
  year = 2018,
  month = 12,
  ico = NULL,
  redownload = FALSE,
  dest_dir = tempdir()
)
```

Arguments

`table_id`	A table ID. See id column in `sp_tables` for a list of available codelists.
`year`	year, numeric, 2015-2018 for some datasets, 2010-2018 for others. Can be a vector of length > 1 (see Details for how to work with data across time periods.).
`month`	month, numeric. Must be 3, 6, 9 or 12. Can be a vector of length > 1 (see details).
`ico`	ID(s) of org to return, character of length one or more. If unset, returns all orgs. ID not checked for correctness/existence. See http://monitor.statnipokladna.cz/datovy-katalog/prohlizec-ciselniku/ucjed to look up ID of any org in the dataset.
`redownload`	Redownload even if recent file present? Defaults to FALSE.
`dest_dir`	character. Directory in which downloaded files will be stored. Defaults to `tempdir()`. Will be created if it does not exist.

Details

Soft-deprecated

Value

a tibble

`sp_add_codelist`	*Add codelist data to downloaded data*

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$period_vykaz,
  by = NULL,
  redownload = FALSE,
  dest_dir = tempdir()
)
```

Arguments

`data`	a data frame returned by `sp_get_table()`.
`codelist`	The codelist to add. Either a character vector of length one (see `sp_tables` for possible values), or a data frame returned by `sp_get_codelist()`.
`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 `sp_get_table()`.
`by`	character. Columns by which to join the codelist. Same form as for dplyr::left_join()``..
`redownload`	Redownload even if file has already been downloaded? Defaults to FALSE.
`dest_dir`	character. Directory in which downloaded files will be stored. Defaults to `tempdir()`.

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 data frame 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

```
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)
```

---

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/ciselniky 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` *Get codelist*

Description

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

Usage

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

Arguments

`codelist_id`	A codelist ID. See id column in `sp_codelists` for a list of available codelists.
`n`	Number of rows to return. Default (NULL) means all. Useful for quickly inspecting a codelist.
`dest_dir`	character. Directory in which downloaded files will be stored. Defaults to `tempdir()`.
`redownload`	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

```
sp_get_codelist("paragraf")
```

---

sp_get_dataset   *Retrieve dataset from statnipokladna*

Description

Downloads and unzips files for a given dataset.

Usage

```
sp_get_dataset(
  dataset_id,
  year = 2018,
  month = 12,
  dest_dir = tempdir(),
  redownload = FALSE
)
```

Arguments

dataset_id	A dataset ID. See id column in sp_datasets for a list of available codelists.
year	year, numeric, 2015-2018 for some datasets, 2010-2019 for others. Defaults to 2018. (see Details for how to work with data across time periods.)
month	month, numeric. 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. Defaults to tempdir(). 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() and further sorted into subdirectories by dataset, year and month. They persist 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 files.

See Also

Other Core workflow: add_codelist(), get_codelist(), sp_add_codelist(), sp_get_codelist(), sp_get_table()

Examples

```
budget_latest <- sp_get_dataset("finm")
budget_2018 <- sp_get_dataset("finm", 2018)
budget_mid2018 <- sp_get_dataset("finm", 2018, 6)
```

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 = ".", download = TRUE)
```

Arguments

dataset_id	dataset ID. See sp_datasets.
dest_dir	character. Where the file should be written. Defaults to working directory.
download	Whether to download (the default) or open link in browser.

Value

character (link) if download = TRUE, nothing otherwise.

See Also

Other Utilities: get_dataset(), sp_get_codelist_viewer()

Examples

```
sp_get_dataset_doc("finm")
```

sp_get_table *Get a statnipokladna table*

Description

Cleans and loads a table. If needed, a dataset containing the table is downloaded.

Usage

```
sp_get_table(
  table_id,
  year = 2018,
  month = 12,
  ico = NULL,
  redownload = FALSE,
  dest_dir = tempdir()
)
```

Arguments

`table_id`	A table ID. See id column in `sp_tables` for a list of available codelists.
`year`	year, numeric, 2015-2018 for some datasets, 2010-2018 for others. Can be a vector of length > 1 (see Details for how to work with data across time periods.).
`month`	month, numeric. Must be 3, 6, 9 or 12. Can be a vector of length > 1 (see details).
`ico`	ID(s) of org to return, character of length one or more. If unset, returns all orgs. ID not checked for correctness/existence. See http://monitor.statnipokladna.cz/datovy-katalog/prohlizec-ciselniku/ucjed to look up ID of any org in the dataset.
`redownload`	Redownload even if recent file present? Defaults to FALSE.
`dest_dir`	character. Directory in which downloaded files will be stored. Defaults to `tempdir()`. Will be created if it does not exist.

Details

The data is loaded from files downloaded automatically by `sp_get_dataset()`; files persist in a temporary directory per session.

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.

Data is processed in the following way:

- all columns are given names that are human-readable and facilitato add codelists
- ICO (org. IDs) are normalised as in some datasets they are padded with leading zeros
- a period, per_yr and per_m columns are created to identify the time period
- value columns are transformed into numeric
- other columns are left as character to avoid losing information

Correspondence between input and output columns:
Shared

Original	Output	English	Czech	Note
ZC_VTAB	vtab	table number	tabulka	-
ZC_UCJED	ucjed	accounting unit	účetní jednotka	NB: ucjed != ico; the two codes are different; ICO is u
ZC_VYKAZ	vykaz	report number	výkaz	-
ZFUNDS_CT	finmisto	accounting centre	finanční místo	either kapitola or "organizační složka státu" (core state
ZC_ICO	ico	org ID	IČO	see ucjed
ZC_FUND	zdroj	funding source	zdroj	-
ZC_KRAJ	kraj	region	kraj	-
ZC_NUTS	nuts	NUTS code	NUTS kód	-

Tables budget-*

Original	Output	English	Czech	Note
ZCMMT_ITM	polozka	item/line	položka (druhové členění)	NB: polo
0FM_AREA	kapitola	chapter	kapitola	-
FUNC0AREA	paragraf	sector line	paragraf (odvětvové členění)	-

0FUNC_AREA	paragraf	sector line	paragraf (odvětvové členění)	-
ZU_ROZSCH	budget_adopted	budget as originally adopted	schválený rozpočet	-
ZU_ROZPZM	budget_amended	budget as amended throughout the year	rozpočet po změnách	-
ZU_KROZP	budget_final	final budget	konečný rozpočet	-
ZU_ROZKZ	budget_spending	actual spending	skutečnost	-

Tables balance-sheet*

Original	Output	English	Czech	Note
ZC_POLVYK	polvyk	item/line	položka výkazu	-
ZU_MONET	previous_net	net, previous period	netto minulé období	-
ZU_AOBTTO	current_gross	gross, current period	brutto běžné období	-
ZU_AONET	current_net	net, current period	netto běžné období	-
ZU_AOKORR	current_correction	correction, current period	korekce běžné období	-

Tables profit-and-loss-*

Original	Output	English	Czech	Note
ZU_HLCIN	previous_core	core activity, previous period	hlavní činnost, minulé období	-
ZU_HOSCIN	previous_economic	economic activity, previous period	hospodářská činnost, minulé období	-
ZU_HLCIBO	current_core	core activity, current period	hlavní činnost, běžné období	-
ZU_HCINBO	current_economic	economic activity, current period	hospodářská činnost, běžné období	-

Table changes-in-equity

Original	Output	English	Czech	Note
ZU_STAVP	before	previous period	stav minulé období	-
ZU_STAVPO	after	current period	stav běžné období	-
ZU_ZVYS	increase	increase	zvýšení stavu	-
ZU_SNIZ	decrease	decrease	snížení stavu	-

Table cash-flow

Original	Output	English	Czech	Note
ZU_BEZUO	current	current period	běžné účetní období	-

Value

a tibble; see Details for key to the columns

See Also

Other Core workflow: add_codelist(), get_codelist(), sp_add_codelist(), sp_get_codelist(), sp_get_dataset()

Examples

```
allorgs_latest <- sp_get_table("budget-local")
```

```
allorgs_2018 <- sp_get_table("budget-local", 2018)
allorgs_mid2018 <- sp_get_table("budget-local", 2018, 6)
oneorg_multiyear <- sp_get_table("budget-local", 2012:2018, 12, ico = "00064581")
oneorg_multihalfyears <- sp_get_table("budget-local", 2013:2018, c(6, 12), ico = "00064581")
```

---

sp_tables   *List of available tables (PARTIAL)*

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/](https://monitor.statnipokladna.cz/datovy-katalog/transakcni-data) [transakcni-data](https://monitor.statnipokladna.cz/datovy-katalog/transakcni-data) to see more detailed descriptions. Note that tables do not correspond to the tab-
ulka/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