Package 'opendataformat'

June 7, 2024

Title Reading and Writing Open Data Format Files

Version 1.1.0

Author Tom Hartl, Claudia Saalbach, Xiaoyao Han, Knut Wenzig

Maintainer Tom Hartl <thartl@diw.de>

Description Reading and writing of data files in the open data format, displaying metadata in different languages.

URL https://git.soep.de/opendata/r-package-opendataformat

BugReports https://git.soep.de/opendata/r-package-opendataformat/-/issues

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Depends R (>= 2.10), crayon, cli, magrittr, xml2, data.table

Suggests testthat (>= 3.0.0), knitr, rmarkdown, devtools, ISLR

VignetteBuilder knitr

VignetteEngine knitr::rmarkdown

VignetteIndexEntry ``opendataformat''

Config/testthat/edition 3

Contents

---

opendataformat-package

Open Data Format

Description

The package is designed to support the use of the open data format. For this purpose, three functions have been developed:

read_opendf():
Import data from the Open Data Format to an R data frame.

write_opendf():
Export data from an R data frame to the open data format.

docu_opendf():
Get access to information about the dataset and variables via the R-Studio Viewer or the web browser.

Author(s)

Claudia Saalbach (<csaalbach@diw.de>), Xiaoyao Han (<xhan@diw.de>), Knut Wenzig (<kwenzig@diw.de>)

Other Contributors: KonsortSWD/NFDI, DIW Berlin

See Also

More information about the Open Data Format specification and data examples are available here: <https://git.soep.de/opendata/>

---

data_opendf   *data_opendf*

Description

example data with attributes specified for the Open Data Format.

Usage

```
data_opendf
```

Format

A data frame with 20 rows and 7 variables:

bap87 Current Health.

bap9201 Hours of sleep, normal workday.

bap9001 Pressed For Time Last 4 Weeks.

bap9002 Run-down, Melancholy Last 4 Weeks.

bap9003 Well-balanced Last 4 Weeks.

bap96 Height.

name Firstname.

Source

https://git.soep.de/opendata/specification

| docu_opendf | *Get documentation from R data frame.* |

Description

Get access to information about the dataset and variables via the R-Studio Viewer or the web browser.

Usage

```
docu_opendf(
  input,
  languages = "current",
  style = "both",
  replace_missing = F,
  variables = "no"
)
```

Arguments

input	R data frame (df) or variable from an R data frame (df$var).
languages	Select the language in which the descriptions and labels of the data will be displayed.

- By default the language that is set to current is displayed (languages = "current").
- The default-option chooses either the default language(if labels and
- descriptions without a language tag exist)Otherwise the current language
- is displayed. (languages = "default").
- You can choose to view all available language variants by selecting (languages = "all"),
- or you can select the language by language code, e.g. languages = "en".

style	Selects where the output should be displayed (console ore viewer).By default the metadata information is displayed in both the console and the viewer. (style = "both") (style = "all")

- You can choose to display the code only in the console (style = "console") (style = "print")
- You can choose to display the code only in the viewer (style = "viewer") (style = "html")

replace_missing	
	If only one language is specified in languages and replace_missings is set to TRUE. In case of a missing label or description, the default or english label/description is displayed additionally (if one of these is available).
variables	Indicate whether a list with all the variables should be displayed with the dataset metadata. If the input is a variable/column, the variables-argument will be ignored. Set (variables="yes") to display the list of variables.

Value

Documentation.

Examples

```
# get example data from the opendataformat package
df <- get(data("data_opendf"))

# view documentation about the dataset in the language that is currently set
## Not run:
docu_opendf(df)

## End(Not run)

# view information from a selected variable in language "en"
## Not run:
docu_opendf(df$bap87, languages = "en")

## End(Not run)

# view dataset information for all available languages
## Not run:
docu_opendf(df, languages = "all")

## End(Not run)

# print information to the R console
## Not run:
docu_opendf(df$bap87, style = "print")

## End(Not run)

# print information to the R viewer
## Not run:
docu_opendf(df$bap87, style = "viewer")

## End(Not run)

# Since the label for language de is missing, in this case the
# english label will be displayed additionally.
## Not run:
attributes(df$bap87)["label_de"]<-""
docu_opendf(df$bap87, languages="de", style = "console", replace_missing=T)

## End(Not run)
```

| labels_opendf | *Get variable labels or other metadata from a data frame in open-dataformat.* |

Description

Get access to information about the dataset and variables via the R-Studio Viewer or the web browser.

Usage

```
labels_opendf(
  input,
  language = "current",
  valuelabels = F,
  retrieve = "labels"
)
```

Arguments

input	R data frame (df) or variable from an R data frame (df$var).
language	Select the language in which the labels of the variables will be displayed. If no language is selected, the current/active language of the data frame will be used.

- By default the language that is set to current is displayed (language = "current").
- You can select the language by language code, e.g. language = "en".

valuelabels	You can choose to display the value labels instead of the variable label, if the input is a variable by setting valuelabels=TRUE or valuelabels="yes". If the input is a dataset, this argument is ignored.
retrieve	You can choose to display another attribute/metadata instead of the variable label(s). Possible options are "description", "url", "type" or "languages".If valuelabels=TRUE or valuelabels="yes" and the input is a variable, this argument is ignored.

Value

Documentation.

Examples

```
# get example data from the opendataformat package
df <- get(data("data_opendf"))
# view the variable labels for all variables in English
labels_opendf(input = df, language = "en", valuelabels = FALSE)

# view the value labels for variable bap87 in English
labels_opendf(input = df$bap87, language = "en", valuelabels = FALSE)

# view the description for variable bap87 in English
labels_opendf(input = df$bap87, language = "en", retrieve = "description")
```

---

read_opendf *Read data specified as Open Data Format.*

Description

Import data from the Open Data Format to an R data frame.

Usage

```
read_opendf(file, languages = "all", nrows = Inf, skip = 0, variables = NULL)
```

Arguments

`file`	the name of the file which the data are to be read from. By default all available language variants are imported (`languages = "all"`).
`languages`	integer: the maximum number of rows to read in. Negative and other invalid values are ignored.
`nrows`	Maximum number of lines to read.
`skip`	Select the number of rows to be skipped (without the column names).
`variables`	Columns to include in the results. You can use the same mini-language as dplyr::select() to refer to the columns by name. Use c() to use more than one selection expression. Although this usage is less common, col_select also accepts a numeric column index. See ?tidyselect::language for full details on the selection language.

Value

R dataframe with attributes including dataset and variable information.

Examples

```
# get path to example data from the opendataformat package (data.zip)
path <- system.file("extdata", "data.zip", package="opendataformat")
path

# read example data specified as Open Data Format from ZIP file
df <- read_opendf(file = path)
df
attributes(df)
attributes(df$bap87)

# read example data with language selection
df <- read_opendf(file = path, languages = "de")
attributes(df$bap87)
```

---

setLanguage_opendf   *Change language of dataframe metadata*

Description

Changes the active language of a dataframe with metadata for the docu_opendf function.

Usage

```
setLanguage_opendf(dataframe, language)
```

Arguments

`dataframe`	R data frame (df) enriched with metadata in the opendf-format.
`language`	Select the language to which you want to switch the metadata.

Value

Dataframe

Examples

```
# get example data from the opendataformat package
df <- get(data("data_opendf"))

# Switch dataset df to language "en"
df <- setLanguage_opendf(df, language = "en")

# Display dataset information for dataset df in language "en"
docu_opendf(df)
```

---

write_opendf   *Write R data frame to the Open Data Format.*

Description

Export data from an R data frame to a ZIP file that stores the data as Open Data Format.

Usage

```
write_opendf(x, file, languages = "all", export_data = TRUE, verbose = T)
```

Arguments

x	R data frame (df) to be writtem.
file	Path to ZIP file or name of zip file to save the opendf-dataset in the working directory.
languages	Select the language in which the descriptions and labels of the data will be exported

  - By default all available language variants are exported (languages = "all").
  - You can also choose to export only the default language (languages = "default"),
  - Or only the current language (languages = "current"),
  - or you can select the language by language code, e.g. languages = "en".

export_data	Choose, if you want to export the file that holds the data (data.csv).Default is TRUE.

  - By default the data and metadata are exported (export_data = "yes").
  - To export only metadata and no data, select export_data = "no"

verbose	Display more messages.

Value

ZIP file and unzipped directory containing the data as CSV file and the metadata as XML file (DDI Codebook 2.5.).

Examples

```
# get example data from the opendataformat package
df <- get(data("data_opendf"))
df

# write R data frame with attributes to the file my_data.zip specified
# as Open Data Format.
## Not run:
write_opendf(x = df, file = "my_data.zip")

## End(Not run)

# write variable from R data frame with attributes to the file my_data.zip
# specified as Open Data Format.
## Not run:
write_opendf(x = df$bap87,  file = "my_data.zip")

## End(Not run)

# write R data frame with attributes to the file my_data.zip
# with selected language.
## Not run:
write_opendf(x = df,  file = "my_data.zip", languages = "en")

## End(Not run)

# write R data frame with attributes to the file my_data.zip but only
# metadata, no data.
## Not run:
write_opendf(x = df,  file = "my_data.zip", export_data = "no")

## End(Not run)
```

Index