

Package ‘febr’

November 12, 2018

Type Package

Title Free Brazilian Repository for Open Soil Data

Version 1.0.1

Date 2018-11-11

Description Making the access to the Free Brazilian Repository for Open Soil Data <<http://www.ufsm.br/febr/>> as easy as possible.

License GPL (>= 2)

Encoding UTF-8

Imports dplyr, glue, googlesheets, knitr, pedometrics, readr, sp, stringr, xlsx

Suggests bookdown, DT, lattice, latticeExtra, magrittr, pander, rgdal

VignetteBuilder knitr

SystemRequirements pandoc

LazyData true

RoxygenNote 6.1.1

URL <https://github.com/febr-team/febr-package/>

BugReports <https://github.com/febr-team/febr-package/issues/>

NeedsCompilation no

Author Alessandro Samuel-Rosa [aut, cre]
(<<https://orcid.org/0000-0003-0877-1320>>),
Universidade Federal de Santa Maria [fnd],
Universidade Tecnológica Federal do Paraná [fnd]

Maintainer Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

Repository CRAN

Date/Publication 2018-11-12 05:40:03 UTC

R topics documented:

dataset	2
febr	3
febr2spdf	4
febr2xlsx	5
goto	5
header	6
layer	7
metadata	10
observation	11
standard	14
unit	15
Index	16

dataset	<i>Get dataset table</i>
---------	--------------------------

Description

Download data from the *dataset* ("dataset") table of one or more datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This includes dataset name and description, author and institution names, dataset license, and much more.

Usage

```
dataset(dataset, progress = TRUE, verbose = TRUE)
```

Arguments

dataset	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use <code>dataset = "all"</code> to download all datasets.
progress	(optional) Logical value indicating if a download progress bar should be displayed.
verbose	(optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to <a href="mailto:<febr-forum@googlegroups.com>"><febr-forum@googlegroups.com> if you find any issue.

Value

A list of data frames or a data frame with data of the chosen dataset(s).

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

Examples

```
res <- dataset(dataset = "ctb0003")
head(res)
```

febr	<i>Get all dataset tables</i>
------	-------------------------------

Description

Download data from the *dataset* ("dataset"), *observation* ("observacao"), *layer* ("camada"), and *metadata* ("metadado") tables of a datasets contained in the Free Brazilian Repository for Open Soil Data – febr, <http://www.ufsm.br/febr>.

Usage

```
febr(dataset, merge = FALSE, progress = TRUE, verbose = TRUE, ...)
```

Arguments

dataset	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use dataset = "all" to download all datasets.
merge	(optional) Logical value indicating if the <i>observation</i> ("observacao") and <i>layer</i> ("camada") tables should be merged. Defaults to merge = FALSE. See merge for more details.
progress	(optional) Logical value indicating if a download progress bar should be displayed.
verbose	(optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to <febr-forum@googlegroups.com> if you find any issue.
...	(optional) Arguments passed to observation and layer .

Value

A list of data frames with data on the chosen dataset.

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

See Also

[dataset](#), [observation](#), [layer](#), [metadata](#)

Examples

```
res <- febr(dataset = "ctb0003", merge = TRUE, variable = "all")
str(res)
```

febr2spdf

Create an object of class SpatialPointsDataFrame

Description

Set spatial coordinates and projection attributes to create an object of class `SpatialPointsDataFrame` from the *observation* ("observacao") table of one or more standardized datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>.

Usage

```
febr2spdf(obj)
```

Arguments

`obj` Object of class `data.frame` downloaded from **febr** using function [observation](#).

Value

An object of class `SpatialPointsDataFrame`

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

Examples

```
library(magrittr)
observation(dataset = "ctb0003", variable = "taxon",
            progress = FALSE, verbose = FALSE) %>%
  febr2spdf() %>%
  sp::splot(zcol = "taxon_sibcs_2009", auto.key = list(columns = 3), scales = list(draw = TRUE))
```

febr2xlsx	<i>Write data to an Excel workbook</i>
-----------	--

Description

Write data downloaded from the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr> – to an Excel workbook.

Usage

```
febr2xlsx(x, file, row.names = FALSE, ...)
```

Arguments

x	A data.frame or list of data.frames to write to the Excel workbook.
file	Character string indicating the path to the output XLSX file.
row.names	(optional) Logical value indicating whether the row names of x are to be written along with x to the file.
...	(optional) Further arguments passed to write.xlsx .

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

Examples

```
dts <-
  febr::febr(dataset = "ctb0013",
            variable = "all",
            merge = TRUE,
            progress = FALSE, verbose = FALSE)
febr2xlsx(x = dts, file = tempfile(fileext = ".xlsx"))
```

goto	<i>Go to febr</i>
------	-------------------

Description

Go to one of the web pages of the Free Brazilian Repository for Open Soil Data – **febr** –, including project and dataset web pages.

Usage

```
goto(dataset, table, page)
```

Arguments

dataset	Character vector indicating one dataset. The identification code should be as recorded in http://www.ufsm.br/febr/catalog/ .
table	Character string indicating a table, i.e. the <i>dataset</i> table, "dataset", the <i>observation</i> table, "observacao", the <i>layer</i> table, "camada", or the <i>metadata</i> table, "metadado".
page	Character string indicating a web page of the febr . Options are: "febr", "view", "catalog", "search", "book", "package", "github", "forum", "unit", and "standard".

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

Examples

```
# Go to the main project page
goto(page = "febr")
```

header	<i>Get table header</i>
--------	-------------------------

Description

Download header data (column names and measurement units) from the *layer* ("camada") or *observation* ("observacao") table of one or more datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This is useful to check what are the variables contained in a dataset before downloading it via [layer](#) or [observation](#).

Usage

```
header(dataset, table, variable, stack = FALSE, progress = TRUE,
        verbose = TRUE)
```

Arguments

dataset	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use dataset = "all" to download all datasets.
table	Character string indicating a table, i.e. the <i>layer</i> table, "camada", or the <i>observation</i> table, "observacao".
variable	(optional) Character vector indicating one or more variables. Accepts only general identification codes, e.g. "ferro" and "carbono". If missing, then a set of standard identification variables is downloaded. Use variable = "all" to download all variables. See 'Details' for more information.

stack	(optional) Logical value indicating if tables from different datasets should be stacked on a single table for output. Requires <code>standardization = list(units = TRUE)</code> – see below. Defaults to <code>stack = FALSE</code> , the output being a list of tables.
progress	(optional) Logical value indicating if a download progress bar should be displayed.
verbose	(optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to febr-forum@googlegroups.com if you find any issue.

Details

Standard identification variables: Standard identification variables and their content depend on the chosen table. See documentation of [layer](#) and [observation](#).

Value

A list of data frames or a data frame with table header data (column names and measurement units) on the chosen variable(s) of the chosen dataset(s).

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

See Also

[layer](#), [observation](#)

Examples

```
res <- header(dataset = c("ctb0001", "ctb0003"), table = "camada", variable = "ferro", stack = TRUE)
id <- grep("ferro_", colnames(res))
col <- colnames(res)[id]
col[order(col)]
```

layer

Get layer table

Description

Download data from the *layer* ("camada") table of one or more datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This includes sampling depth, horizon designation, and variables such as pH, carbon content, clay content, and much more. Use [header](#) if you want to check what are the variables contained in the *layer* table of a dataset before downloading it.

Usage

```
layer(dataset, variable, stack = FALSE, missing = list(depth = "keep",
  data = "keep"), standardization = list(plus.sign = "keep", plus.depth =
  2.5, lessthan.sign = "keep", lessthan.frac = 0.5, repetition = "keep",
  combine.fun = "mean", transition = "keep", smoothing.fun = "mean", units
  = FALSE, round = FALSE), harmonization = list(harmonize = FALSE, level
  = 2), progress = TRUE, verbose = TRUE)
```

Arguments

dataset	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use dataset = "all" to download all datasets.
variable	(optional) Character vector indicating one or more variables. Accepts only general identification codes, e.g. "ferro" and "carbono". If missing, then a set of standard identification variables is downloaded. Use variable = "all" to download all variables. See 'Details' for more information.
stack	(optional) Logical value indicating if tables from different datasets should be stacked on a single table for output. Requires standardization = list(units = TRUE) – see below. Defaults to stack = FALSE, the output being a list of tables.
missing	(optional) List with named sub-arguments indicating what should be done with a layer missing data on sampling depth, depth, or data on variable(s), data. Options are "keep" (default) and "drop".
standardization	(optional) List with named sub-arguments indicating how to perform data standardization. <ul style="list-style-type: none"> • plus.sign Character string indicating what should be done with the plus sign (+) commonly used along with the inferior limit of the bottom layer of an observation. Options are "keep" (default), "add", and "remove". • plus.depth Numeric value indicating the depth increment (in centimeters) when processing the plus sign (+) with plus.sign = "add". Defaults to plus.depth = 2.5. • lessthan.sign Character string indicating what should be done with the less-than sign (<) used to indicate that the value of a variable is below the lower limit of detection. Options are "keep" (default), "subtract", and "remove". • lessthan.frac Numeric value between 0 and 1 (a fraction) by which the lower limit of detection should be subtracted when lessthan.sign = "subtract". Defaults to lessthan.frac = 0.5, i.e. subtract 50% from the lower limit of detection. • repetition Character string indicating what should be done with repetitions, i.e. repeated measurements of layers in an observation. Options are "keep" (default) and "combine". In the latter case, it is recommended to set lessthan.sign = "subtract" or lessthan.sign = "remove". • combine.fun Character string indicating the function that should be used to combine repeated measurements of layers in an observation when repetition = "combine". Options are "mean" (default), "min", "max", and "median".

- **transition** Character string indicating what should be done about the wavy and irregular transition between subsequent layers in an observation. Options are "keep" (default) and "smooth".
 - **smoothing.fun** Character string indicating the function that should be used to smooth wavy and irregular transitions between subsequent layers in an observation when `transition = "smooth"`. Options are "mean" (default), "min", "max", and "median".
 - **units** Logical value indicating if the measurement units of the continuous variable(s) should be converted to the standard measurement unit(s). Defaults to `units = FALSE`, i.e. no conversion is performed. See [standard](#) for more information.
 - **round** Logical value indicating if the values of the continuous variable(s) should be rounded to the standard number of decimal places. Requires `units = TRUE`. Defaults to `round = FALSE`, i.e. no rounding is performed. See [standard](#) for more information.
- harmonization** (optional) List with named sub-arguments indicating if and how to perform data harmonization.
- **harmonize** Logical value indicating if data should be harmonized? Defaults to `harmonize = FALSE`, i.e. no harmonization is performed.
 - **level** Integer value indicating the number of levels of the identification code of the variable(s) that should be considered for harmonization. Defaults to `level = 2`. See 'Details' for more information.
- progress** (optional) Logical value indicating if a download progress bar should be displayed.
- verbose** (optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to <febr-forum@googlegroups.com> if you find any issue.

Details

Standard identification variables: Standard identification variables and their content are as follows:

- `dataset_id`. Identification code of the dataset in **febr** to which an observation belongs.
- `observacao_id`. Identification code of an observation in **febr**.
- `camada_id`. Sequential layer number, from top to bottom.
- `camada_nome`. Layer designation according to some standard description guide.
- `amostra_id`. Laboratory number of a sample.
- `profund_sup`. Upper boundary of a layer (cm).
- `profund_inf`. Lower boundary of a layer (cm).

Further details about the content of the standard identification variables can be found in <http://www.ufsm.br/febr/book/> (in Portuguese).

Harmonization: Data harmonization consists of converting the values of a variable determined using some method *B* so that they are (approximately) equivalent to the values that would have been obtained if the standard method *A* had been used instead. For example, converting carbon

content values obtained using a wet digestion method to the standard dry combustion method is data harmonization.

A heuristic data harmonization procedure is implemented in the **febr** package. It consists of grouping variables based on a chosen number of levels of their identification code. For example, consider a variable with an identification code composed of four levels, `aaa_bbb_ccc_ddd`, where `aaa` is the first level and `ddd` is the fourth level. Now consider a related variable, `aaa_bbb_eee_fff`. If the harmonization is to consider all four coding levels (`level = 4`), then these two variables will remain coded as separate variables. But if `level = 2`, then both variables will be re-coded as `aaa_bbb`, thus becoming the same variable.

Value

A list of data frames or a data frame with data on the chosen variable(s) of the chosen dataset(s).

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

See Also

[observation](#), [standard](#), [unit](#)

Examples

```
res <- layer(dataset = "ctb0013")
str(res)
```

metadata

Get metadata table

Description

Download data from the *metadata* ("metadado") table of one or more datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This includes variable names, description of analytical methods, and identification of analysis laboratories.

Usage

```
metadata(dataset, progress = TRUE, verbose = TRUE)
```

Arguments

dataset	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use <code>dataset = "all"</code> to download all datasets.
progress	(optional) Logical value indicating if a download progress bar should be displayed.

`verbose` (optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to <febr-forum@googlegroups.com> if you find any issue.

Value

A list of data frames or a data frame with metadata of the chosen dataset(s).

Author(s)

Alessandro Samuel-Rosa [<alessandrosamuelrosa@gmail.com>](mailto:alessandrosamuelrosa@gmail.com)

Examples

```
res <- metadata(dataset = "ctb0003")
head(res)
```

observation	<i>Get observation table</i>
-------------	------------------------------

Description

Download data from the *observation* ("observacao") table of one or more datasets contained in the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This includes spatial coordinates, observation date, and variables such as geology, land use and vegetation, local topography, and much more. Use [header](#) if you want to check what are the variables contained in the *observation* table of a dataset before downloading it.

Usage

```
observation(dataset, variable, stack = FALSE, missing = list(coord =
  "keep", time = "keep", data = "keep"), standardization = list(crs =
  NULL, time.format = NULL, units = FALSE, round = FALSE),
  harmonization = list(harmonize = FALSE, level = 2), progress = TRUE,
  verbose = TRUE)
```

Arguments

<code>dataset</code>	Character vector indicating one or more datasets. Identification codes should be as recorded in http://www.ufsm.br/febr/catalog/ . Use <code>dataset = "all"</code> to download all datasets.
<code>variable</code>	(optional) Character vector indicating one or more variables. Accepts only general identification codes, e.g. "ferro" and "carbono". If missing, then a set of standard identification variables is downloaded. Use <code>variable = "all"</code> to download all variables. See 'Details' for more information.
<code>stack</code>	(optional) Logical value indicating if tables from different datasets should be stacked on a single table for output. Requires <code>standardization = list(units = TRUE)</code> – see below. Defaults to <code>stack = FALSE</code> , the output being a list of tables.

missing	(optional) List with named sub-arguments indicating what should be done with an observation missing spatial coordinates, coord, date of observation, time, or data on variables, data? Options are "keep" (default) and "drop".
standardization	(optional) List with named sub-arguments indicating how to perform data standardization. <ul style="list-style-type: none"> • <code>crs</code> Character string indicating the EPSG code of the coordinate reference system (CRS) to which spatial coordinates should be transformed. For example, <code>crs = "EPSG:4674"</code>, i.e. SIRGAS 2000, the standard CRS for Brazil – see more at http://spatialreference.org/ref/epsg/. Defaults to <code>crs = NULL</code>, i.e. no transformation is performed. • <code>time.format</code> Character string indicating how to format dates. For example, <code>time.format = "%d-%m-%Y"</code>, i.e. dd-mm-yyyy such as in 31-12-2001. Defaults to <code>time.format = NULL</code>, i.e. no formatting is performed. See as.Date for more details. • <code>units</code> Logical value indicating if the measurement units of the continuous variable(s) should be converted to the standard measurement unit(s). Defaults to <code>units = FALSE</code>, i.e. no conversion is performed. See standard for more information. • <code>round</code> Logical value indicating if the values of the continuous variable(s) should be rounded to the standard number of decimal places. Requires <code>units = TRUE</code>. Defaults to <code>round = FALSE</code>, i.e. no rounding is performed. See standard for more information.
harmonization	(optional) List with named sub-arguments indicating if and how to perform data harmonization. <ul style="list-style-type: none"> • <code>harmonize</code> Logical value indicating if data should be harmonized? Defaults to <code>harmonize = FALSE</code>, i.e. no harmonization is performed. • <code>level</code> Integer value indicating the number of levels of the identification code of the variable(s) that should be considered for harmonization. Defaults to <code>level = 2</code>. See 'Details' for more information.
progress	(optional) Logical value indicating if a download progress bar should be displayed.
verbose	(optional) Logical value indicating if informative messages should be displayed. Generally useful to identify datasets with inconsistent data. Please report to febr-forum@googlegroups.com if you find any issue.

Details

Standard identification variables: Standard identification variables and their content are as follows:

- `dataset_id`. Identification code of the dataset in **febr** to which an observation belongs.
- `observacao_id`. Identification code of an observation in **febr**.
- `sisb_id`. Identification code of an observation in the Brazilian Soil Information System maintained by the Brazilian Agricultural Research Corporation (EMBRAPA) at https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html.

- `ibge_id`. Identification code of an observation in the database of the Brazilian Institute of Geography and Statistics (IBGE) at http://www.downloads.ibge.gov.br/downloads_geociencias.htm#.
- `observacao_data`. Date (dd-mm-yyyy) in which an observation was made.
- `coord_sistema`. EPSG code of the coordinate reference system.
- `coord_x`. Longitude (°) or easting (m).
- `coord_y`. Latitude (°) or northing (m).
- `coord_precisao`. Precision with which x- and y-coordinates were determined (m).
- `coord_fonte`. Source of the x- and y-coordinates.
- `pais_id`. Country code (ISO 3166-1 alpha-2).
- `estado_id`. Code of the Brazilian federative unit where an observation was made.
- `municipio_id`. Name of the Brazilian county where as observation was made.
- `amostra_tipo`. Type of sample taken.
- `amostra_quant`. Number of samples taken.
- `amostra_area`. Sampling area.

Further details about the content of the standard identification variables can be found in <http://www.ufsm.br/febr/book/> (in Portuguese).

Harmonization: Data harmonization consists of converting the values of a variable determined using some method *B* so that they are (approximately) equivalent to the values that would have been obtained if the standard method *A* had been used instead. For example, converting carbon content values obtained using a wet digestion method to the standard dry combustion method is data harmonization.

A heuristic data harmonization procedure is implemented in the **febr** package. It consists of grouping variables based on a chosen number of levels of their identification code. For example, consider a variable with an identification code composed of four levels, `aaa_bbb_ccc_ddd`, where `aaa` is the first level and `ddd` is the fourth level. Now consider a related variable, `aaa_bbb_eee_fff`. If the harmonization is to consider all four coding levels (`level = 4`), then these two variables will remain coded as separate variables. But if `level = 2`, then both variables will be re-coded as `aaa_bbb`, thus becoming the same variable.

Value

A list of data frames or a data frame with data on the chosen variable(s) of the chosen dataset(s).

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

See Also

[layer](#), [standard](#), [unit](#)

Examples

```
res <- observation(dataset = "ctb0013", variable = "taxon")
str(res)
```

 standard

Get standard table

Description

Download data from the standard ("febr-padrao") table of the Free Brazilian Repository for Open Soil Data

febr, <http://www.ufsm.br/febr>. This includes codes and names of variable, measurement units, number of decimal places, type of data, and description of analytical methods. This is used to standardize the data contained in a dataset when downloading it via [layer](#) or [observation](#).

Usage

```
standard(table, variable, unit, precision, expr)
```

Arguments

table	(optional) Character string indicating a table, i.e. the <i>layer</i> table, "camada", or the <i>observation</i> table, "observacao".
variable	(optional) Character vector indicating one or more variables. Accepts both specific identification codes, e.g. "ferro_oxalato_icpoes" and "carbono_cromo_30min150_mohr", as well as general identification codes, e.g. "ferro" and "carbono".
unit	(optional) Character vector indicating one or more measurement units. For example, "g/kg", "g/cm^3", and "cmolc/kg".
precision	(optional) Integer vector indicating one or more number of decimal places.
expr	(optional) Character string to be parsed and evaluated as a regular expression. For example, "campo_precisao > 0". Usage requires some knowledge of the structure of the <i>standard</i> table.

Value

A data frame with standards for selected variable(s).

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

References

Teixeira, P. C., Donagemma, G. K., Fontana, A., Teixeira, W. G. (2017) *Manual de Métodos de Análise de Solo*. Brasília: Embrapa.

See Also

The *standard* table at <https://goo.gl/hi77sB>

Examples

```
res <- standard(variable = "ferro")
head(res)
```

unit

Get unit table

Description

Download data from the *unit* ("unidade") table of the Free Brazilian Repository for Open Soil Data – **febr**, <http://www.ufsm.br/febr>. This includes measurement units and conversion factors. This is used to standardize the data contained in a dataset when downloading it via [layer](#) or [observation](#).

Usage

```
unit(source, target)
```

Arguments

source (optional) Character vector indicating one or more source measurement units.
target (optional) Character vector indicating one or more target measurement units.

Value

A data.frame with source and target measurement units and their corresponding conversion factors.

Author(s)

Alessandro Samuel-Rosa <alessandrosamuelrosa@gmail.com>

References

Teixeira, P. C., Donagemma, G. K., Fontana, A., Teixeira, W. G. (2017) *Manual de Métodos de Análise de Solo*. Brasília: Embrapa.

See Also

The *unit* table at <https://goo.gl/Vvvsf2>

Examples

```
res <- unit(source = c("%", "dag/kg"), target = "g/kg")
res
```

Index

`as.Date`, [12](#)

`dataset`, [2](#), [3](#)

`febr`, [3](#)

`febr2spdf`, [4](#)

`febr2xlsx`, [5](#)

`goto`, [5](#)

`header`, [6](#), [7](#), [11](#)

`layer`, [3](#), [6](#), [7](#), [7](#), [13–15](#)

`merge`, [3](#)

`metadata`, [3](#), [10](#)

`observation`, [3](#), [4](#), [6](#), [7](#), [10](#), [11](#), [14](#), [15](#)

`standard`, [9](#), [10](#), [12](#), [13](#), [14](#)

`unit`, [10](#), [13](#), [15](#)

`write.xlsx`, [5](#)