

Package: odkr (via r-universe)

September 12, 2024

Type Package

Title 'Open Data Kit' ('ODK') R API

Version 0.3.3.9000

Description Utility functions for working with datasets gathered using 'Open Data Kit' ('ODK') <<https://opendatakit.org/>>. These include an API to interface with 'ODK Briefcase', a 'Java' application for fetching and pushing 'ODK' forms and their contents, that allows pulling of data from either a remote 'ODK Aggregate Server' or a local 'ODK' folder, a rename function to give more human readable variable names for 'ODK' datasets, a merge function to create a single dataframe from a nested 'ODK' dataset and an expand function to disaggregate multiple choice answers that have been collapsed into single code by 'ODK'.

License GPL-3

Depends R (>= 2.10)

SystemRequirements Java (>= 8)

Imports stringr, Hmisc, curl, rJava

Suggests testthat (>= 3.0.0), spelling, covr

Encoding UTF-8

Language en-GB

LazyData true

RoxygenNote 7.3.1

Roxygen list(markdown = TRUE)

URL <https://github.com/rapidsurveys/odkr>,
<https://rapidsurveys.io/odkr>,

BugReports <https://github.com/rapidsurveys/odkr/issues>

Config/testthat/edition 3

Repository <https://rapidsurveys.r-universe.dev>

RemoteUrl <https://github.com/rapidsurveys/odkr>

RemoteRef HEAD

RemoteSha b581df1891755ddeb22cf8e6bb4c48f765f48a05

Contents

create_sd	2
expandMultChoice	3
expand_choice	4
export_data	5
get_briefcase	7
get_help	7
mergeNestedODK	8
pull_local	9
pull_remote	10
push_data	12
renameODK	13
sampleData1	14
sampleData2	15
sampleData3	16
Index	17

create_sd	<i>Function to create an ODK Briefcase Storage directory</i>
-----------	--

Description

Function to create an ODK Briefcase Storage directory

Usage

```
create_sd(path)
```

Arguments

path Directory path on which to create the ODK Briefcase Storage

Value

ODK Briefcase Storage directory in the specified path

Examples

```
create_sd(path = tempdir())
```

expandMultChoice	<i>Function to recode a choose-multiple column where multiple answers are recorded in the same element (e.g. "ab ac ad", "ef eg eh") to a data frame with multiple separate 0/1 columns e.g.:</i>
------------------	---

Description

ab	ac	ad	ef	eg	eh
1	1	1	0	0	0
0	0	0	1	1	1

Usage

```
expandMultChoice(answers, choices = NULL, naCode = NULL, naQuestion = NULL)
```

Arguments

answers	Character vector with given answers (strings containing the choices)
choices	(optional) Character vector with choices to be used (each will become a column). If not supplied, choices will be determined from the answer parameter, splitting given answers by an empty space (" ") and finding unique values.
naCode	(optional) Single element specifying what character code equates to NA
naQuestion	(optional) TRUE/FALSE vector of the same length as answers; in rows where this is false, all columns will be coded as NA

Value

a data frame with multiple separate 0/1 columns

NOTE

naCode must exist as the only answer in a column (an answer that contains both a valid answer and the NA code will not be recognized as NA - instead, the NA code will be output as an extra answer column (if choices parameter is not given))

Examples

```
# Expand responses in variable w7 of sampleData2
sampleData2 <- renameODK(sampleData2)
temp <- expandMultChoice(sampleData2$w7)
```

expand_choice	<i>Function to expand response to a more than one answer multiple choice question coded as a concatenated string</i>
---------------	--

Description

Function to expand response to a more than one answer multiple choice question coded as a concatenated string

Usage

```
expand_choice(
  df,
  x,
  values,
  pattern = "",
  prefix = x,
  labels = values,
  sep = "."
)
```

Arguments

df	A dataframe containing the vector data that requires expansion.
x	Name of variable in df containing vector needing expansion.
values	Vector of string values used to create concatenated string response.
pattern	Pattern used to separate values in the concatenated string. Default is "" for concatenated strings with no separator.
prefix	Prefix to names of newly created variables.
labels	Vector of names to use for columns of resulting data.frame. If not specified, columns are named using values.
sep	Character to separate prefix from labels in the names of newly created variables.

Value

A data.frame with same rows as df containing columns corresponding to each newly created variable.

Examples

```
## Not run:
expandMultipleChoice(df = individual,
  x = "mddw1",
  values = as.character(0:18),
  pattern = " ")
```

```

        prefix = "mddw",
        sep = "")

## End(Not run)

```

export_data	<i>Export data in CSV format from local ODK Briefcase Storage directory to a specified destination directory and a specified file name</i>
-------------	--

Description

Export data in CSV format from local ODK Briefcase Storage directory to a specified destination directory and a specified file name

Usage

```

export_data(
  target = "",
  briefcase = "odkBriefcase_latest",
  sd = FALSE,
  id = "",
  from = "",
  to = "",
  filename = paste(id, "_data.csv", sep = ""),
  start = NULL,
  end = NULL,
  overwrite = FALSE,
  exclude = TRUE,
  group.names = TRUE,
  split = FALSE,
  pem = NULL,
  pullBefore = FALSE,
  includeGeo = FALSE
)

```

Arguments

target	Path to directory of ODK Briefcase .jar file. Directory path should match directory path used when calling <code>get_briefcase()</code> . If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , target should match the directory path where .jar file has been downloaded into.
briefcase	Filename of the downloaded ODK Briefcase .jar file. Default is <code>odkBriefcase_latest</code> to match the default filename used by <code>get_briefcase()</code> . If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , filename should match the default filename used by Open Data Kit which is usually "ODK Briefcase vX.Y.Z Production.jar" where vX.Y.Z is the version number

sd	Logical. If TRUE, create an ODK Briefcase Storage in the path specified by to. Default is FALSE.
id	Form ID of form to be pulled
from	Path to source ODK Briefcase Storage from which to extract data. This should match directory path specified when making a call to pull_remote() or pull_local()
to	Destination directory to save output data file
filename	Filename of output CSV data; default is paste(id, "_data.csv", sep = "")
start	Include data from submission dates after (inclusive) this start date in export to CSV. Date format YYYY-MM-DD or YYYY/MM/DD
end	Include data from submission dates before (inclusive) this date in export to CSV. Date format YYYY-MM-DD> or YYYY/MM/DD
overwrite	Overwrite existing output data in destination directory with the same filename; default is FALSE
exclude	Exclude media files on export; default is TRUE
group.names	Logical. Should group names be removed from column names on export? Default TRUE.
split	Logical. Should select multiple fields be split on export? Default FALSE.
pem	Path to pem key if using an encrypted form. Null by default.
pullBefore	Logical. If set to true, pull before export. Default FALSE.
includeGeo	Logical. If set to true, pull geojson. Default FALSE.

Value

CSV file in destination directory containing data from the pulled forms

Examples

```
# Export data from a specified ODK Briefcase Storage directory to current
# working directory with a filename called "test.csv"
## Not run:
  dirPath <- tempdir()
  get_briefcase(destination = dirPath)
  export_data(target = dirPath,
             from = dirPath,
             to = dirPath,
             id = "stakeholders",
             filename = "test.csv",
             overwrite = TRUE)

## End(Not run)
```

get_briefcase	<i>Updates pre-installed ODK Briefcase jar file to the latest version downloaded from https://opendatakit.org.</i>
---------------	--

Description

Updates pre-installed ODK Briefcase jar file to the latest version downloaded from <https://opendatakit.org>.

Usage

```
get_briefcase(destination = "", briefcase = "odkBriefcase_latest")
```

Arguments

destination	Path to directory where ODK Briefcase .jar file will be downloaded into
briefcase	Filename of the downloaded ODK Briefcase .jar file. Default is "odkBriefcase_latest"

Examples

```
# Get latest version of ODK Briefcase and save in a temporary directory
## Not run:
  dirPath <- tempdir()
  get_briefcase(destination = dirPath)

## End(Not run)
```

get_help	<i>Get help with command line interface for ODK Briefcase</i>
----------	---

Description

Get help with command line interface for ODK Briefcase

Usage

```
get_help(target = "", briefcase = "odkBriefcase_latest")
```

Arguments

target	Path to directory of ODK Briefcase .jar file. Directory path should match directory path used when calling get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , target should match the directory path where .jar file has been downloaded into.
--------	--

briefcase Filename of the downloaded ODK Briefcase .jar file. Default is `odkBriefcase_latest` to match the default filename used by `get_briefcase()`. If ODK Briefcase .jar file was downloaded manually from <https://opendatakit.org>, filename should match the default filename used by Open Data Kit which is usually "ODK Briefcase vX.Y.Z Production.jar" where vX.Y.Z is the version number

Value

Help notes on usage of ODK Briefcase via command line interface

Examples

```
## Not run:
dirPath <- tempdir()
get_briefcase(destination = dirPath)
get_help(target = dirPath)

## End(Not run)
```

<code>mergeNestedODK</code>	<i>Retrieve parent data, matching to a nested (child) dataset by keys of a nested dataset exported from an ODK Aggregate Server.</i>
-----------------------------	--

Description

Retrieve parent data, matching to a nested (child) dataset by keys of a nested dataset exported from an ODK Aggregate Server.

Usage

```
mergeNestedODK(
  parent,
  child,
  byPARENT_KEY = TRUE,
  removeCols = NULL,
  removeRows = NULL
)
```

Arguments

<code>parent</code>	Data frame of household data
<code>child</code>	Data frame of child (repeat) data
<code>byPARENT_KEY</code>	Should data frames be matched based on PARENT_KEY (child) and KEY (parent) columns? (Currently only option is TRUE)
<code>removeCols</code>	Character vector of column names to remove from the parent data frame (optional)
<code>removeRows</code>	Index of which rows should be removed from child data frame (optional)

Value

Merged dataframe

Examples

```
# merge sampleData2 and sampleData3
x <- renameODK(sampleData2)
y <- renameODK(sampleData3)
temp <- mergeNestedODK(parent = x, child = y, byPARENT_KEY = FALSE)
```

pull_local	<i>Pull ODK forms from a local ODK folder (/odk) collected from ODK Collect mobile clients</i>
------------	--

Description

Pull ODK forms from a local ODK folder (/odk) collected from ODK Collect mobile clients

Usage

```
pull_local(
  target = "",
  briefcase = "odkBriefcase_latest",
  id = "",
  to = "",
  from = "",
  pem = NULL
)
```

Arguments

target	Path to directory of ODK Briefcase .jar file. Directory path should match directory path used when calling get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , target should match the directory path where .jar file has been downloaded into.
briefcase	Filename of the downloaded ODK Briefcase .jar file. Default is odkBriefcase_latest to match the default filename used by get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , filename should match the default filename used by Open Data Kit which is usually "ODK Briefcase vX.Y.Z Production.jar" where vX.Y.Z is the version number
id	Form ID of form to be pulled
to	Destination directory for pulled ODK forms
from	Source ODK directory (/odk) from ODK Collect mobile client
pem	If form to be pulled is encrypted, a PEM private key file would be required to pull forms; default is NULL; if form is encrypted, provide path to PEM file

Value

Folder in destination directory named ODK Briefcase Storage containing forms pulled from local ODK folder

Examples

```
# Pull forms from a local ODK folder to current working directory
## Not run:
dirPath <- tempdir()
get_briefcase(destination = dirPath)
pull_local(target = dirPath,
           id = "stakeholders",
           from = system.file("odk", package = "odkr"),
           to = dirPath)

## End(Not run)
```

pull_remote

Pull ODK forms from remote ODK Aggregate via ODK Briefcase

Description

Pull ODK forms from remote ODK Aggregate via ODK Briefcase

Usage

```
pull_remote(
  target = "",
  briefcase = "odkBriefcase_latest",
  sd = FALSE,
  id = "",
  to = "",
  from = "",
  include_incomplete = FALSE,
  max_http_connections = NULL,
  username,
  password
)
```

Arguments

target Path to directory of ODK Briefcase .jar file. Directory path should match directory path used when calling get_briefcase(). If ODK Briefcase .jar file was downloaded manually from <https://opendatakit.org>, target should match the directory path where .jar file has been downloaded into.

briefcase	Filename of the downloaded ODK Briefcase .jar file. Default is odkBriefcase_latest to match the default filename used by get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , filename should match the default filename used by Open Data Kit which is usually "ODK Briefcase vX.Y.Z Production.jar" where vX.Y.Z is the version number
sd	Logical. If TRUE, create an ODK Briefcase Storage in the path specified by to. Default is FALSE.
id	Form ID of form to be pulled
to	Destination directory for pulled ODK forms
from	URL of remote ODK Aggregate server to pull ODK forms data from
include_incomplete	Logical. Should incomplete forms be pulled? Default to FALSE
max_http_connections	Integer value for maximum simultaneous HTTP connections allowed. Defaults to NULL which will allow for the default 8 simultaneous HTTP connections. Specify this parameter if more simultaneous connections are required. Maximum value is 32.
username	Username for account in remote ODK Aggregate server from which forms are to be pulled
password	Password for account in remote ODK Aggregate server from which forms are to be pulled

Value

Folder in destination directory named "ODK Briefcase Storage" containing forms pulled from remote ODK Aggregate server

Examples

```
# Use latest ODK Briefcase and connect to a test
# remote ODK Aggregate server from ONA (https://ona.io); pulled forms to
# be saved in default location at current working directory
## Not run:
dirPath <- tempdir()
get_briefcase(destination = dirPath)
pull_remote(target = dirPath,
            id = "stakeholders",
            from = "https://ona.io/validtrial",
            to = dirPath,
            username = "validtrial",
            password = "zEF-STN-5ze-qom")

## End(Not run)
```

push_data	<i>Push ODK forms from local ODK Briefcase Storage folder to remote ODK Aggregate via ODK Briefcase</i>
-----------	---

Description

Push ODK forms from local ODK Briefcase Storage folder to remote ODK Aggregate via ODK Briefcase

Usage

```
push_data(
    target = "",
    briefcase = "odkBriefcase_latest",
    id = "",
    to = "",
    from = "",
    force_send_blank = FALSE,
    max_http_connections = NULL,
    username,
    password
)
```

Arguments

target	Path to directory of ODK Briefcase .jar file. Directory path should match directory path used when calling get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , target should match the directory path where .jar file has been downloaded into.
briefcase	Filename of the downloaded ODK Briefcase .jar file. Default is odkBriefcase_latest to match the default filename used by get_briefcase(). If ODK Briefcase .jar file was downloaded manually from https://opendatakit.org , filename should match the default filename used by Open Data Kit which is usually "ODK Briefcase vX.Y.Z Production.jar" where vX.Y.Z is the version number
id	Form ID of form to push ODK forms data into
to	URL of remote ODK Aggregate server
from	Directory containing ODK forms data to push to remote ODK aggregate server
force_send_blank	Logical. Should blank form be forced into the Aggregate instance
max_http_connections	Integer value for maximum simultaneous HTTP connections allowed. Defaults to NULL which will allow for the default 8 simultaneous HTTP connections. Specify this parameter if more simultaneous connection required
username	Username for account in remote ODK Aggregate server from which forms are to be pulled

password Password for account in remote ODK Aggregate server from which forms are to be pulled

Examples

```
# Use latest ODK Briefcase and connect to a test
# remote ODK Aggregate server from ONA (https://ona.io) to push ODK forms
# data into
## Not run:
dirPath <- tempdir()
get_briefcase(destination = dirPath)
push_data(target = dirPath,
          id = "stakeholders",
          to = "https://ona.io/validtrial",
          from = dirPath,
          username = "validtrial",
          password = "zEF-STN-5ze-qom")

## End(Not run)
```

renameODK	<i>Rename column names of data exported from an ODK Aggregate Server or from ODK Briefcase into more usable and human readable variable names.</i>
-----------	--

Description

Rename column names of data exported from an ODK Aggregate Server or from ODK Briefcase into more usable and human readable variable names.

Usage

```
renameODK(data, sep = c(".", "-"))
```

Arguments

data	Dataframe object of dataset exported from ODK Aggregate Server or from local ODK directory
sep	Character value for separator used in variable names. Choices are "." or "-". Default is ".".

Value

Data frame object with renamed variables

Examples

```
# Rename sampleData1 dataset to remove '.' from variable names
names(sampleData1)
renameODK(sampleData1)
names(sampleData1)

# Rename sampleData2 dataset
names(sampleData2)
renameODK(sampleData2)
names(sampleData2)

# Rename sampleData3 dataset
names(sampleData3)
renameODK(sampleData3)
names(sampleData3)
```

sampleData1

Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains cluster level data from the survey.

Description

Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains cluster level data from the survey.

Usage

```
sampleData1
```

Format

A data frame with 16 columns and 50 rows:

```
admin.admin1.adm1 Date of survey
admin.admin1.adm2 Enumerator ID
admin.enameA Name of enumerator (Arabic)
admin.ename Name of enumerator (English)
admin.admin2.adm3 Survey round number
admin.admin2.adm4 Study area / cluster
location.loc1 Village ID
location.loc1a Village ID - other
location.loc2 Is village a replacement village
location.loc3 Village population
```

location.loc4 GPS coordinates
 hh1 Household ID
 hh2 Number of women aged 15-49 in household
 wcount_count Current woman respondent's ID
 KEY Parent data identifier
 PARENT_KEY Child data identifier

Examples

sampleData1

sampleData2	<i>Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains information from mother respondents.</i>
-------------	---

Description

Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains information from mother respondents.

Usage

sampleData2

Format

A data frame with 16 columns and 50 rows:

wcount.wdata.women.wage Mother's age
 wcount.wdata.women.wmarried Is mother married?
 wcount.wdata.women.wpregnant Is mother pregnant?
 wcount.wdata.women.wedu1 Mother - number of years of formal education
 wcount.wdata.women.wedu2 Mother - highest educational attainment
 wcount.wdata.women.wanthro Mother's middle upper arm circumference (mm)
 wcount.wdata.women.screening Has mother's MUAC and oedema been measured/tested in past month
 wcount.wdata.wash.ws1 Source of drinking water
 wcount.wdata.wash.ws2 Water treatment
 wcount.wdata.wash.ws3 Sanitation facility
 wcount.wdata.wash.ws4 Is sanitation facility shared with other households?
 wcount.wdata.wash.ws5 Is sanitation facility used by public

wcount.wdata.wash.ws6 Sanitary disposal of child's faeces
wcount.wdata.wash.ws7 Episodes when handwashing is done
KEY Parent data identifier
PARENT_KEY Child data identifier

Examples

sampleData2

sampleData3	<i>Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains information from child respondents.</i>
-------------	--

Description

Sample dataset from an impact evaluation study of a mother and child nutrition programme in Kassala State, Sudan. This dataset contains information from child respondents.

Usage

sampleData3

Format

A data frame with 9 columns and 50 rows:

wcount.wdata.ccount.child.csex Child's gender
wcount.wdata.ccount.child.card Does child have an immunisation card?
wcount.wdata.ccount.child.cdob Child's date of birth
wcount.wdata.ccount.child.cage Age of child
wcount.wdata.ccount.illness.ill1 Has child had diarrhoea in the past 2 weeks
wcount.wdata.ccount.illness.ill2 Has child had fever in the past 2 weeks
wcount.wdata.ccount.illness.ill3 Has child had cough in the past 2 weeks
KEY Parent data identifier
PARENT_KEY Child data identifier

Examples

sampleData3

Index

* datasets

sampleData1, [14](#)

sampleData2, [15](#)

sampleData3, [16](#)

create_sd, [2](#)

expand_choice, [4](#)

expandMultChoice, [3](#)

export_data, [5](#)

get_briefcase, [7](#)

get_help, [7](#)

mergeNestedODK, [8](#)

pull_local, [9](#)

pull_remote, [10](#)

push_data, [12](#)

renameODK, [13](#)

sampleData1, [14](#)

sampleData2, [15](#)

sampleData3, [16](#)