

Package ‘rapidpror’

December 3, 2024

Title Read Data From a RapidPro API

Version 0.0.1

Description Convenience functions to help reading data from a RapidPro API into R.

License LGPL (>= 3) + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Suggests testthat (>= 2.0.0)

Config/testthat/edition 2

Imports stringr,

dplyr,

magrittr,

plyr,

purrr,

jsonlite,

tibble,

tidyr,

httr,

utils

Depends R (>= 2.10)

Contents

days_active_data	2
filter_date_range	3
flow_data_calculation	4
get_archived_data	5
get_data_from_rapidpro_api	6
get_flow_data	7
get_flow_names	8
get_rapidpro_key	9
get_rapidpro_site	10
get_rapidpro_uuid_names	10
get_user_data	10
handle_type_flow	12
httr_get_call	13

link_data_frames	13
set_rapidpro_key	14
set_rapidpro_site	14
set_rapidpro_uuid_names	15
unlist_rapidpro_variable	15
update_archived_data	16

Index	17
--------------	-----------

days_active_data	<i>Count number of active days for each user</i>
------------------	--

Description

The number of days that each user is active

Usage

```
days_active_data(
  uuid_data = get_rapidpro_uuid_names(),
  flow_name,
  rapidpro_site = get_rapidpro_site(),
  token = get_rapidpro_key(),
  flatten = FALSE,
  read_runs = FALSE,
  runs_data = "result_flow_runs.RDS",
  include_archived_data = FALSE,
  get_by = "gotit",
  data_from_archived = archived_data,
  read_archived_data_from = "archived_data_monthly.RDS"
)
```

Arguments

uuid_data	A string containing UUID values. See set_rapidpro_uuid_names() to set this value.
flow_name	A string containing flow names to call data from.
rapidpro_site	A string containing the rapidpro website to call the data from. See set_rapidpro_site() to amend the website.
token	A string containing the token to call the data. See set_rapidpro_key() to amend the token.
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
read_runs	Default FALSE. A boolean denoting whether to read in the runs_data, or to read the data from RapidPro
runs_data	If the "runs.json" data has been read in already, this is the run data. Instead, the uuid_data, call_type, and token can be given (note that this latter approach is slower since it requires reading the data in).
include_archived_data	Default FALSE. A boolean denoting whether to include archived data or not.

`get_by` A string denoting how to access the archived data (if `include_archived_data` is `TRUE`). Options are `"read_from_RDS"`, `"gotit"`. The first denotes that the data is read from an RDS file (given in `read_archived_data_from`), the latter that the data is stored and will be read in directly (see `data_from_archived`).

`data_from_archived`
If `get_by = "gotit"`, this is the archived data.

`read_archived_data_from`
If `get_by = "read_from_RDS"`, this is the archived data.

Value

A data frame with two columns: The unique ID of the individual, and the number of days they have been active.

<code>filter_date_range</code>	<i>Helper function to filter data by date range.</i>
--------------------------------	--

Description

This function filters a data frame based on a specified date range.

Usage

```
filter_date_range(data, date, format, tzzone, start = TRUE)
```

Arguments

<code>data</code>	A data frame to be filtered.
<code>date</code>	The date to filter the data by.
<code>format</code>	The date-time format as used by <code>strptime</code> .
<code>tzzone</code>	The time zone specification to be used for conversion.
<code>start</code>	A logical value indicating whether to filter for dates before or after the specified date.

Value

A filtered data frame.

flow_data_calculation *Calculate flow data.*

Description

This function arranges flow data for different flow types.

Usage

```
flow_data_calculation(
  result_flow,
  flatten = FALSE,
  flow_type = "none",
  flow_handle_type = NULL,
  flow_handle_type_sub = "category",
  date_from = NULL,
  date_to = NULL,
  format_date = "%Y-%m-%d",
  tzzone_date = "UTC",
  created_on = FALSE
)
```

Arguments

result_flow	Data frame containing flow data.
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
flow_type	Default none, option of other to specify a handle type.
flow_handle_type	Default NULL. If flow_type is other, this is the value of the flow under \$value to get the response for.
flow_handle_type_sub	Default NULL. If flow_type is other, this is the group within flow_handle_type to handle.
date_from	character string giving the date to filter the data from.
date_to	character string giving the date to filter the data to.
format_date	character string giving a date-time format as used by <code>strptime</code> .
tzzone_date	time zone specification to be used for the conversion, if one is required.
created_on	Parameter for flow_data.

Value

Data frame displaying the flow data.

get_archived_data	<i>Get archived (flow/run) data</i>
-------------------	-------------------------------------

Description

Download the archived run data from RapidPro.

Usage

```
get_archived_data(
    rapidpro_site = get_rapidpro_site(),
    token = get_rapidpro_key(),
    period = "monthly",
    flatten = FALSE,
    date_from = NULL,
    date_to = NULL,
    format_date = "%Y-%m-%d",
    tzone_date = "UTC"
)
```

Arguments

rapidpro_site	A string containing the rapidpro website to call the data from. See <code>set_rapidpro_site()</code> to amend the website.
token	A string containing the token to call the data. See <code>set_rapidpro_key()</code> to amend the token.
period	Default "monthly". A character string giving the period of time. One of "monthly", "daily", or "none".
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
date_from	character string giving the date to filter the data from.
date_to	character string giving the date to filter the data to.
format_date	from <code>as.POSIX*</code> function: character string giving a date-time format as used by <code>strptime</code> .
tzone_date	from <code>as.POSIX*</code> function: time zone specification to be used for the conversion, if one is required. System-specific (see time zones), but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Value

Archived data

get_data_from_rapidpro_api

Calling data from RapidPro API

Description

A generalised way to call data from the RapidPro API.

Usage

```
get_data_from_rapidpro_api(
  call_type,
  rapidpro_site = get_rapidpro_site(),
  token = get_rapidpro_key(),
  flatten = FALSE,
  date_from = NULL,
  date_to = NULL,
  format_date = "%Y-%m-%d",
  tzone_date = "UTC"
)
```

Arguments

call_type	A string containing the call type; for example, "flows.json".
rapidpro_site	A string containing the rapidpro website to call the data from. See set_rapidpro_site() to amend the website.
token	A string containing the token to call the data. See set_rapidpro_key() to amend the token.
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
date_from	character string giving the date to filter the data from.
date_to	character string giving the date to filter the data to.
format_date	from as.POSIX* function: character string giving a date-time format as used by strptime.
tzone_date	from as.POSIX* function: time zone specification to be used for the conversion, if one is required. System-specific (see time zones), but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Value

A data frame of the data specified in the call_type parameter.

get_flow_data	<i>Get data from each run (flow)</i>
---------------	--------------------------------------

Description

Call RapidPro data from each run or flow.

Usage

```
get_flow_data(
  uuid_data = get_rapidpro_uuid_names(),
  flow_name = NULL,
  rapidpro_site = get_rapidpro_site(),
  token = get_rapidpro_key(),
  by = c("flow_name", "id_name"),
  id_names = NULL,
  flatten = FALSE,
  checks = FALSE,
  return_all = FALSE,
  flow_type = "none",
  flow_handle_type = NULL,
  flow_handle_type_sub = "category",
  include_archived_data = FALSE,
  get_by = "gotit",
  data_from_archived = NULL,
  read_archived_data_from = NULL,
  created_on = FALSE,
  date_from = NULL,
  date_to = NULL,
  format_date = "%Y-%m-%d",
  tzzone_date = "UTC"
)
```

Arguments

uuid_data	A string containing UUID values. See <code>set_rapidpro_uuid_names()</code> to set this value.
flow_name	A string containing flow names to call data from if <code>by = "flow_name"</code> .
rapidpro_site	A string containing the rapidpro website to call the data from. See <code>set_rapidpro_site()</code> to amend the website.
token	A string containing the token to call the data. See <code>set_rapidpro_key()</code> to amend the token.
by	A string taking either <code>"flow_name"</code> or <code>"id_name"</code> to determine whether to filter the flow data by a given flow name, or to give all flows for a given id.
id_names	A string containing ID names to call data from if <code>by = "id_name"</code> .
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
checks	Default FALSE. A boolean whether to check and update the data if it is not found.

return_all	Default FALSE. A boolean whether to return a specific part of the string, or all of the data.
flow_type	Only if return_all = FALSE. Default none. Takes values "other".
flow_handle_type	Default NULL. If flow_type is other, this is the value of the flow under \$value to get the response for.
flow_handle_type_sub	Default NULL. If flow_type is other, this is the group within flow_handle_type to handle.
include_archived_data	Default FALSE. A boolean denoting whether to include archived data or not.
get_by	A string denoting how to access the archived data (if include_archived_data is TRUE). Options are "read_from_RDS", "gotit". The first denotes that the data is read from an RDS file (given in read_archived_data_from), the latter that the data is stored and will be read in directly (see data_from_archived).
data_from_archived	If get_by = "gotit", this is the archived data.
read_archived_data_from	If get_by = "read_from_RDS", this is the archived data.
created_on	Parameter for flow_data.
date_from	character string giving the date to filter the data from.
date_to	character string giving the date to filter the data to.
format_date	from as.POSIX* function: character string giving a date-time format as used by strptime.
tzzone_date	from as.POSIX* function: time zone specification to be used for the conversion, if one is required. System-specific (see time zones), but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Value

List separated by each flow_name provided. Each element in the list contains a data frame for each flow_name provided.

get_flow_names	<i>Get a data frame containing the UUID and names for flows.</i>
----------------	--

Description

Call the names of each run (flow) and their UUID values. This function is often called inside another function.

Usage

```

get_flow_names(
  rapidpro_site = get_rapidpro_site(),
  token = get_rapidpro_key(),
  flatten = FALSE,
  date_from = NULL,
  date_to = NULL,
  format_date = "%Y-%m-%d",
  tzone_date = "UTC"
)

```

Arguments

rapidpro_site	A string containing the rapidpro website to call the data from. See set_rapidpro_site() to amend the website.
token	A string containing the token to call the data. See set_rapidpro_key() to amend the token.
flatten	Default FALSE. A boolean denoting whether the data should be flattened into a two-dimensional tabular structure.
date_from	character string giving the date to filter the data from.
date_to	character string giving the date to filter the data to.
format_date	from as.POSIX* function: character string giving a date-time format as used by strptime.
tzone_date	from as.POSIX* function: time zone specification to be used for the conversion, if one is required. System-specific (see time zones), but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Value

A data frame containing the flows data in RapidPro.

get_rapidpro_key	<i>Get the key (token) from the environment</i>
------------------	---

Description

Call the key (token) from the environment. The key is set in the function set_rapidpro_key.

Usage

```
get_rapidpro_key()
```

Value

returns a string containing the key (Token) set in the environment for the rapidPro data.

get_rapidpro_site	<i>Get the website from the environment</i>
-------------------	---

Description

Call the website from the environment. The website is set in the function set_rapidpro_site.

Usage

```
get_rapidpro_site()
```

Value

returns a string containing the website name set in the environment for the RapidPro data.

get_rapidpro_uuid_names	<i>Get the UUID names from the environment</i>
-------------------------	--

Description

Call the UUID names from the environment. This set in the function set_rapidpro_uuid_names.

Usage

```
get_rapidpro_uuid_names()
```

Value

returns the UUID names set in the environment for the RapidPro data.

get_user_data	<i>Get data for each user from RapidPro</i>
---------------	---

Description

This function retrieves contact data from RapidPro for each user. It is a wrapper function around get_data_from_rapidpro_api and allows for filtering and formatting of the returned data.

Usage

```

get_user_data(
  rapidpro_site = get_rapidpro_site(),
  token = get_rapidpro_key(),
  call_type = "contacts.json",
  filter_variable = "group",
  filter_variable_value = "joined",
  flatten = FALSE,
  date_from = NULL,
  date_to = NULL,
  format_date = "%Y-%m-%d",
  tzone_date = "UTC",
  unlist_consent = FALSE
)

```

Arguments

rapidpro_site	A string containing the RapidPro website URL from which to retrieve the data. Use <code>set_rapidpro_site()</code> to amend the website.
token	A string containing the API token required for authentication. Use <code>set_rapidpro_key()</code> to set the token.
call_type	A string specifying the type of data call. The default is "contacts.json". To get user data for a specific group, such as "joined", set this parameter to "contacts.json?group=joined", or, for multiple groups use the <code>filter_variable</code> and <code>filter_variable_value</code> parameters.
filter_variable	A string specifying the variable used to filter the data. The default is "group". This variable is appended to the <code>call_type</code> to form the final API endpoint.
filter_variable_value	A string specifying the value of <code>filter_variable</code> used to filter the data. For example, "joined" can be used to filter contacts by group.
flatten	Logical, default is FALSE. If TRUE, the data will be flattened into a two-dimensional tabular structure.
date_from	A character string representing the start date to filter the data from. The date format should match <code>format_date</code> .
date_to	A character string representing the end date to filter the data to. The date format should match <code>format_date</code> .
format_date	A character string specifying the date-time format, as used by <code>strptime</code> . The default is "%Y-%m-%d".
tzone_date	A character string specifying the time zone for date conversion. The default is "UTC". This parameter is passed to <code>as.POSIXct</code> to ensure correct time zone handling.
unlist_consent	Logical, default is FALSE. If TRUE, any variables containing "consent" in their name will be unlisted, simplifying the data structure.

Details

This function builds the API call to retrieve user data from RapidPro, with optional filtering by group and date. The retrieved data can be flattened into a simple data frame, and optional date filters can be applied. If `unlist_consent` is set to TRUE, variables containing "consent" will be unlisted to simplify the structure.

Value

A data frame containing the contact data from RapidPro, filtered and formatted according to the specified parameters.

Examples

```
#\dontrun{
## Set the RapidPro site and API key
#set_rapidpro_site("https://app.rapidpro.io")
#set_rapidpro_key("your_api_key")
#
# # Get data for users in the "joined" group
# user_data <- get_user_data(call_type = "contacts.json", filter_variable = "group", filter_variable_value = "
#
# # Get data for users in the "joined" group, filtering by date range
# user_data <- get_user_data(call_type = "contacts.json", filter_variable = "group", filter_variable_value = "
# }
```

handle_type_flow

Helper function to handle flow data of a specified type.

Description

This function processes flow data of a specified type and extracts relevant information.

Usage

```
handle_type_flow(data, type = "completed", type_2 = "category")
```

Arguments

data	A data frame containing flow data.
type	The type of flow data to handle.
type_2	The group within type to handle.

Value

A tibble with relevant information.

httr_get_call	<i>Calling data from RapidPro</i>
---------------	-----------------------------------

Description

Call data from RapidPro into R. The main purpose of this function is to be called within other functions.

Usage

```
httr_get_call(get_command, token = get_rapidpro_key())
```

Arguments

get_command	A string containing the website to call the data from. Usually called within a function.
token	A string containing the token to call the data. See set_rapidpro_key() to amend the token.

Value

A data frame containing RapidPro data.

link_data_frames	<i>Link Data Frames within R-Instat</i>
------------------	---

Description

This function links two data frames within the R-Instat environment using specified linking pairs. It is specifically designed for use within R-Instat and will stop with an error message if the data book object is not found, indicating that the data frames are not being managed through R-Instat.

Usage

```
link_data_frames(
  from_data_frame = "flow_data",
  to_data_frame = "user_data",
  link_pairs = c(uuid = "uuid")
)
```

Arguments

from_data_frame	The name of the source data frame to link from, defaults to "flow_data".
to_data_frame	The name of the target data frame to link to, defaults to "user_data".
link_pairs	A named vector that specifies the column names used to establish the link, with the name being the column in the from_data_frame and the value being the corresponding column in the to_data_frame. Defaults to linking "uuid" in both data frames.

Value

None; the function modifies the data_book environment by adding a new link.

Examples

```
# Assuming 'data_book' and required data frames are already loaded in R-Instat:
# link_data_frames("survey_data", "participant_data", c(participant_id="id"))
```

set_rapidpro_key	<i>Reset the key value in the environment</i>
------------------	---

Description

Set a key (token) value to access the given RapidPro website.

Usage

```
set_rapidpro_key(key, file = FALSE)
```

Arguments

key	A string denoting the API Token associated with the website given in set_rapidpro_site. If file = FALSE, the string is the key value itself. If file = TRUE, the string is a text file containing the key. Only the first line will be taken in the text file.
file	A boolean denoting whether the key given is a file path or a string.

Value

A string containing the key (token) required to access the given RapidPro website.

set_rapidpro_site	<i>Reset the website in the environment</i>
-------------------	---

Description

Reset the website in the environment

Usage

```
set_rapidpro_site(site)
```

Arguments

site	A string denoting the website associated with the token given in set_rapidpro_key.
------	--

Value

A string containing the website required to access RapidPro.

Examples

```
#
```

`set_rapidpro_uuid_names`*Set the UUID names in the environment for the flow data*

Description

Set the UUID names to access data in RapidPro.

Usage

```
set_rapidpro_uuid_names(uuid_names = get_flow_names())
```

Arguments

<code>uuid_names</code>	Default <code>get_flow_names()</code> (recommended). Data frame containing the flow names and UUIDs.
-------------------------	--

Value

Setting a data frame containing RapidPro flow data to the environment.

`unlist_rapidpro_variable`*Unlist variables in a variable*

Description

After importing data from RapidPro, unlist the variables that are stored in a list

Usage

```
unlist_rapidpro_variable(data, variable = "groups")
```

Arguments

<code>data</code>	The data frame containing the data
<code>variable</code>	The variable that contains the groups

Value

returns a data frame containing the unlisted data.

update_archived_data	<i>Update archived data</i>
----------------------	-----------------------------

Description

Read in current archived data, and update it.

Usage

```
update_archived_data(  
    curr_data,  
    period = "none",  
    date_to = NULL,  
    format_date = "%Y-%m-%d",  
    tzone_date = "UTC"  
)
```

Arguments

curr_data	Archived data set from RapidPro (can be generated by get_archived_data).
period	Default "none". A character string giving the period of time. One of "monthly", "daily", or "none".
date_to	character string giving the date to filter the data to.
format_date	from as.POSIX* function: character string giving a date-time format as used by strptime.
tzone_date	from as.POSIX* function: time zone specification to be used for the conversion, if one is required. System-specific (see time zones), but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Value

Updated archived data

Index

days_active_data, [2](#)

filter_date_range, [3](#)
flow_data_calculation, [4](#)

get_archived_data, [5](#)
get_data_from_rapidpro_api, [6](#)
get_flow_data, [7](#)
get_flow_names, [8](#)
get_rapidpro_key, [9](#)
get_rapidpro_site, [10](#)
get_rapidpro_uuid_names, [10](#)
get_user_data, [10](#)

handle_type_flow, [12](#)
http_get_call, [13](#)

link_data_frames, [13](#)

set_rapidpro_key, [14](#)
set_rapidpro_site, [14](#)
set_rapidpro_uuid_names, [15](#)

unlist_rapidpro_variable, [15](#)
update_archived_data, [16](#)