

Package: rgeopi (via r-universe)

January 17, 2025

Type Package

Title Gather Data from GDOT's GeoPI Database

Version 0.1.0

Description rGeoPI allows the user to easily interface with GDOT's GeoPI repository. This includes project information, phase years and costs, project geometry, and document information, returned in data frames.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports dplyr, glue, httr, httr2, lubridate, magrittr, memoise, polite, purrr, rlang, rvest, sf, stringr, tibble, tidyr, tidyselect, vctrs

RoxygenNote 7.3.2

Depends R (>= 2.10)

URL <https://github.com/ModernMobilityPartners/rgeopi>

BugReports <https://github.com/ModernMobilityPartners/rgeopi/issues>

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Config/pak/sysreqs git make libgit2-dev libicu-dev libxml2-dev libssl-dev libudunits2-dev libx11-dev

Repository <https://ar-puuk.r-universe.dev>

RemoteUrl <https://github.com/ModernMobilityPartners/rgeopi>

RemoteRef HEAD

RemoteSha 4b2e7630d526caf964c86bca7e316dc6fbc100fe

Contents

check_pi	2
check_rtxt	3
geopi_session	3
get_geopi	4
get_geopi_docs	5
get_geopi_overview	6
get_geopi_phase	7
get_geopi_sf	8
guess_basename	9
polite_download_file	9
polite_fetch_rtxt	10
polite_read_html	11
%otherwise%	11

Index	13
--------------	-----------

check_pi	<i>Check if PI is possible based on format</i>
----------	--

Description

GDOT PIs are seven character IDs formatted as 1) numbers with leading zeros, 2) six numbers and a "-", and 3) six numbers with a leading M or S. Checks the format to skip running a GeoPI call if the project ID is not possible for the system.

Usage

```
check_pi(gdot_pi)
```

Arguments

gdot_pi	GDOT PI. If not seven digits, will be padded if it's numeric
---------	--

Value

True/False

Examples

```
check_pi("0012345")
check_pi("546540-")
check_pi("M023424")
check_pi("FAKE523")
```

check_rtxt	<i>Function for checking robots.txt file</i>
------------	--

Description

Function for checking robots.txt file

Usage

```
check_rtxt(url, delay, user_agent, force, verbose)
```

Arguments

url	web address for download
delay	default delay
user_agent	user agent string
force	force re-downloading of robots.txt
verbose	logical

geopi_session	<i>geopi session</i>
---------------	----------------------

Description

'geopi_session' masks 'polite::bow', creating a polite session with GeoPI. This can be plugged into any of the 'get_geopi' family of functions.

Usage

```
geopi_session(...)
```

Arguments

... any arguments passed to 'polite::bow' can be passed to

Value

object of class 'polite', 'session'

Examples

```
session <- geopi_session()  
session
```

get_geopi

Get GeoPI Data

Description

Get the available GeoPI data for a project. Adds the field 'Gather.Date' with the current date to document the date of extraction.

The function 'get_geopi_ef' is a more efficient version of 'get_geopi' that avoids pinging the GeoPi website excess times. It is a more recent development and may eventually replace 'get_geopi'.

Usage

```
get_geopi(
  gdot_pi,
  session = NULL,
  features = c("overview", "phases", "documents"),
  doc_mode = c("cr_only", "cr_check", "doc_summary"),
  geometry = FALSE,
  pi_check = TRUE,
  gather_date = NULL
)
```

```
get_geopi_ef(
  gdot_pi,
  session = NULL,
  features = c("overview", "phases", "documents"),
  doc_mode = c("cr_only", "cr_check", "doc_summary"),
  geometry = FALSE,
  pi_check = TRUE,
  gather_date = NULL
)
```

Arguments

gdot_pi	GDOT Project ID. Can take a list or vector of IDs.
session	If NULL (the default), creates a new session. Can provide a session made with 'polite::bow'. In line with polite/ethical scraping, you can use an existing session or the function will create one for you.
features	Project features desired to retrieve. Can choose between "overview" (project name, description, etc.), "phases" (phases, their years, and money allocated), and "documents" (information about what documents GeoPI has for the project).
doc_mode	If "documents" is chosen, the doc_mode conveys what information to retrieve. Options are "cr_only" (description of all files under "approved concept reports"), "cr_check" (simple TRUE/FALSE for if the project has approved concept reports), and "doc_summary" (the name, file path, and type of all project documents).

geometry	if FALSE (the default), do not return spatial data. if TRUE, uses the 'get_geopi_sf' function to add a sf tibble named "geometry" to the output list.
pi_check	Check if the PI is a valid format. Defaults to TRUE.
gather_date	Date information is gathered from GeoPI. Defaults to today.

Details

rgeopi can only access the first 50 phase records. Having more than 50 phases for a single project is rare.

Value

a list of tibbles

Examples

```
## Not run:
get_geopi(gdot_pi = "0000820", doc_mode = "cr_check")

## End(Not run)
```

get_geopi_docs	<i>Check for Concept Reports</i>
----------------	----------------------------------

Description

Check for Concept Reports

Usage

```
get_geopi_docs(
  gdot_pi,
  session = NULL,
  mode = c("cr_only", "cr_check", "doc_summary"),
  gather_date = NULL,
  pi_check = TRUE
)

get_geopi_docs_ef(
  gdot_pi,
  page_scrape,
  dcmode = c("cr_only", "cr_check", "doc_summary"),
  gather_date = NULL
)
```

Arguments

gdot_pi	GDOT Project ID. Can take a list or vector of IDs.
session	If NULL (the default), creates a new session. Can provide a session made with 'polite::bow'.
mode	if "cr_only" (the default), returns the filepath and filenames of all approved concept reports. If "cr_check", merely returns T/F for each project if it has a concept report. If "doc_summary", returns the file name, type, and path for all project documents.
gather_date	Date information is gathered from GeoPI. Defaults to today.
pi_check	Check if the PI is a valid format. Defaults to TRUE.
page_scrape	Scraped GeoPI Page supplied by 'get_geopi_ef'. Only present in 'get_geopi_docs_ef'.
dcmode	if "cr_only" (the default), returns the filepath and filenames of all approved concept reports. If "cr_check", merely returns T/F for each project if it has a concept report. If "doc_summary", returns the file name, type, and path for all project documents.

Value

a tibble

Examples

```
## Not run:
get_geopi_docs(gdot_pi = "0000820", mode = "cr_check")

## End(Not run)
```

get_geopi_overview *Get Project Overview*

Description

Get Project Overview

Usage

```
get_geopi_overview(
  gdot_pi,
  session = NULL,
  gather_date = NULL,
  pi_check = TRUE
)

get_geopi_overview_ef(gdot_pi, page_scrape, gather_date = NULL)
```

Arguments

gdot_pi	GDOT Project ID. Can take a list or vector of IDs.
session	If NULL (the default), creates a new session. Can provide a session made with 'polite::bow'.
gather_date	Date information is gathered from GeoPI. Defaults to today.
pi_check	Check if the PI is a valid format. Defaults to TRUE.
page_scrape	Scraped GeoPI Page supplied by 'get_geopi_ef'. Only present in 'get_geopi_overview_ef'.

Value

a tibble

Examples

```
## Not run:
get_geopi_overview(gdot_pi = "0000820")

## End(Not run)
```

get_geopi_phase	<i>Get Project Phase</i>
-----------------	--------------------------

Description

Get the Phase ID, Programmed Year, Date of Last Estimate, and Cost Estimate for a Project ID.

Usage

```
get_geopi_phase(gdot_pi, session = NULL, gather_date = NULL, pi_check = TRUE)

get_geopi_phase_ef(gdot_pi, page_scrape, gather_date = NULL)
```

Arguments

gdot_pi	GDOT Project ID. Can take a list or vector of IDs.
session	If NULL (the default), creates a new session. Can provide a session made with 'polite::bow'.
gather_date	Date information is gathered from GeoPI. Defaults to today.
pi_check	Check if the PI is a valid format. Defaults to TRUE.
page_scrape	Scraped GeoPI Page supplied by 'get_geopi_ef'. Only present in 'get_geopi_phase_ef'.

Details

rgeopi can only access the first 50 phase records. Having more than 50 phases for a single project is rare. If the project has more than 50 records, the column "Surplus.Records" is added with the number of total records and pages.

Value

a tibble

Examples

```
## Not run:  
get_geopi_phase(gdot_pi = "0000820")  
  
## End(Not run)
```

get_geopi_sf	<i>Get GeoPI Spatial Data</i>
--------------	-------------------------------

Description

Get the spatial geometry of a GDOT project. 'get_geopi_sf' uses the httr2 package.

Usage

```
get_geopi_sf(gdot_pi, pi_check = TRUE)
```

Arguments

gdot_pi	GDOT Project ID. Can be a single PI or a vector or list of PIs. If more than one PI is provided, each project will get its own row.
pi_check	Check if the PI is a valid format. Defaults to TRUE.

Value

An sf tibble

Examples

```
## Not run:  
get_geopi_sf(gdot_pi = "0000820")  
  
## End(Not run)
```

guess_basename	<i>Guess filename for download from url</i>
----------------	---

Description

Guess filename for download from url

Usage

```
guess_basename(x)
```

Arguments

x url to guess filename from

Value

guessed file name

polite_download_file	<i>Polite download</i>
----------------------	------------------------

Description

Polite download

Usage

```
polite_download_file(  
  url,  
  destfile = guess_basename(url),  
  ...,  
  quiet = !verbose,  
  mode = "wb",  
  path = "downloads/",  
  user_agent = paste0("polite ", getOption("HTTPUserAgent")),  
  delay = 5,  
  force = FALSE,  
  overwrite = FALSE,  
  verbose = FALSE  
)
```

Arguments

url	web address for the file to be downloaded
destfile	name of destination file
...	additional arguments passed to 'download.file'
quiet	default value is inverse of 'verbose'
mode	download mode. Default value is "wb"
path	path to save. Default path 'downloads/'
user_agent	default value 'paste0("polite ", getOption("HTTPUserAgent"))'
delay	default value equal 5
force	force re-download of robots.txt
overwrite	overwrite downloaded file. Default value FALSE
verbose	default value is FALSE

Value

An (invisible) integer code, 0 for success and non-zero for failure.

polite_fetch_rtxt *Function to get robots.txt is structured form. Memoised*

Description

Function to get robots.txt is structured form. Memoised

Usage

```
polite_fetch_rtxt(..., user_agent, delay, verbose)
```

Arguments

...	arguments passed to 'robotstxt::robotstxt()'
user_agent	user agent string
delay	default delay
verbose	logical

polite_read_html *function that actually fetches response from the web*

Description

function that actually fetches response from the web

Usage

```
polite_read_html(
  url,
  ...,
  delay = 5,
  user_agent = paste0("polite ", getOption("HTTPUserAgent"), "bot"),
  force = FALSE,
  verbose = FALSE
)
```

Arguments

url	web address for scraping
...	arguments passed to 'httr::GET()'
delay	scraping delay. Default 5 sec
user_agent	user agent string. Default value 'paste0("polite ", getOption("HTTPUserAgent"), "bot")'
force	force re-download of robots.txt
verbose	default FALSE

%otherwise% *null-coalescing operator. See purr for details.*

Description

null-coalescing operator. See purr for details.

Usage

```
lhs %otherwise% rhs
```

Arguments

lhs	left hand side
rhs	right hand side

Examples

```
## Not run:  
a %otherwise% b  
  
## End(Not run)
```

Index

`%otherwise%`, [11](#)

`check_pi`, [2](#)

`check_rtxt`, [3](#)

`geopi_session`, [3](#)

`get_geopi`, [4](#)

`get_geopi_docs`, [5](#)

`get_geopi_docs_ef (get_geopi_docs)`, [5](#)

`get_geopi_ef (get_geopi)`, [4](#)

`get_geopi_overview`, [6](#)

`get_geopi_overview_ef (get_geopi_overview)`, [6](#)

`get_geopi_phase`, [7](#)

`get_geopi_phase_ef (get_geopi_phase)`, [7](#)

`get_geopi_sf`, [8](#)

`guess_basename`, [9](#)

`polite_download_file`, [9](#)

`polite_fetch_rtxt`, [10](#)

`polite_read_html`, [11](#)