

Package: rcmip6 (via r-universe)

October 16, 2024

Title Download CMIP6 Data

Version 0.0.2.9000

Description Search and download CMIP6 data.

License GPL-3

URL <https://github.com/eliocamp/rcmip6>

BugReports <https://github.com/eliocamp/rcmip6/issues>

Encoding UTF-8

Language en

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Imports glue, jsonlite, httr, digest, data.table, curl, furr,
progress, ps

Suggests rstudioapi, testthat (>= 3.0.0), covr

Config/testthat/edition 3

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

RemoteUrl <https://github.com/eliocamp/rcmip6>

RemoteRef HEAD

RemoteSha 0a417ed902bfb93c2df0d0cdfc4a0b6b2147d4eb

Contents

cmip_available	2
cmip_download	2
cmip_filter_replicas	3
cmip_root_set	4
cmip_search	4
cmip_simplify	5
cmip_size	6
cmip_urls	6

Index	7
--------------	----------

cmip_available	<i>List downloaded models.</i>
----------------	--------------------------------

Description

Reads files in your local CMIP6 directory and parses the contents into a data.frame

Usage

```
cmip_available(root = cmip_root_get())  
cmip_available_legacy(root = cmip_root_get())
```

Arguments

root Root folder to download and organise the data.

Details

cmip_available_legacy() reads from the older way of storing information and will only work for files downloaded with version $\leq 0.0.2$ of the package.

Value

A data.frame.

cmip_download	<i>Downloads CMIP data</i>
---------------	----------------------------

Description

Downloads CMIP data

Usage

```
cmip_download(  
  results,  
  root = cmip_root_get(),  
  user = Sys.info()[["user"]],  
  comment = NULL,  
  year_range = c(-Inf, Inf),  
  check_diskspace = TRUE,  
  download_config = cmip_download_config(),  
  ...  
)
```

```

cmip_download_config(
  delay = 0.5,
  retry = 5,
  total_connections = 1,
  host_connections = 1,
  low_speed_limit = 100,
  low_speed_time = 30
)

```

Arguments

results	A list of search results from <code>cmip_search()</code> .
root	Root folder to download and organise the data.
user, comment	Deprecated.
year_range	An integer vector of length 2, indicating the start and end range of years. Restricts the download of model output with files that include some data within this range of years. Defaults to <code>c(-Inf, Inf)</code> to include all possible files
check_diskspace	Logical indicating whether to check if location has enough space to download all the requested files.
download_config	a list of arguments to configure the behaviour of downloads.
...	Ignored
delay	delay in seconds between retries. The actual delay adds a bit of randomness.
retry	number of retries before giving up on a download.
total_connections	maximum number of total concurrent connections.
host_connections	maximum number concurrent connections per host.
low_speed_limit, low_speed_time	the download will fail if it downloads at below <code>low_speed_limit</code> bytes/second for more than <code>low_speed_time</code> seconds.

Value

A list of files.

`cmip_filter_replicas` *Filters replicas based on online servers*

Description

Checks the online status of data nodes and returns filtered results so that there is a single replica of each instance that is hosted on an online node.

Usage

```
cmip_filter_replicas(results)
```

Arguments

results A list of search results from `cmip_search()`.

Value

A data.table.

cmip_root_set	<i>Set or get the root folder for CMIP data.</i>
---------------	--

Description

Set or get the root folder for CMIP data.

Usage

```
cmip_root_set(root, mode = "default")
```

```
cmip_root_get()
```

Arguments

root	Root folder.
mode	Mode that defines folder permissions. Can be an umask valid string or the aliases: <ul style="list-style-type: none"> • "default": the default umask returned by <code>Sys.umask(NA)</code>. • "shared": read and write permissions to any user (mode "0000") • "private": read and write permissions only for the current user (mode "7777")

cmip_search	<i>Query CMIP data</i>
-------------	------------------------

Description

Query CMIP data

Usage

```
cmip_search(query)
```

```
cmip_info(results)
```

```
cmip_url_to_list(url)
```

Arguments

query	A list that defines the search parameters. Optionally, a character vector of instances IDs to be searched.
results	A list of search results from <code>cmip_search()</code> .
url	The URL of the JSON query to convert into a list. See details.

Details

The easiest way to get a valid query list is to use the search portal at (<https://esgf-node.llnl.gov/search/cmip6/>) to make a search that approximates what you are looking for. Then, under the number of results there's a link that reads "return results as JSON". Copy that link and pass it to `[cmip_url_to_list()]`. On RStudio you can also use the AddIn.

Value

A data.table of results.

cmip_simplify	<i>Simplifies the output of searches</i>
---------------	--

Description

Removes a lot of less important columns from the output of `cmip_search()`. The full dataset is stored in the hidden column `full_info`. Use `cmip_unsimplify()` to return to the original format

Usage

```
cmip_simplify(results)
```

```
cmip_unsimplify(data)
```

Arguments

results	A list of search results from <code>cmip_search()</code> .
data	A simplified output from <code>cmip_simplify()</code>

cmip_size	<i>Computes the total size of a search result in Mb.</i>
-----------	--

Description

Computes the total size of a search result in Mb.

Usage

```
cmip_size(results)
```

Arguments

results	A list of search results from cmip_search() .
---------	---

cmip_urls	<i>Obtain file server URLs of CMIP data</i>
-----------	---

Description

Finds the URLs that [cmip_download](#) would obtain from a search results object.

Usage

```
cmip_urls(results)
```

Arguments

results	A list of search results from cmip_search() .
---------	---

Details

If a valid URL is not found the return value has an NA in its place.

Note that the versions to download have "/fileServer/" in the path, whereas those for the remote access have "/dodsC/" instead (and can be string-substituted). This works for some sources we have seen (GDAL for example can within limitations read remotely from the 'fileServer' URLs, but the NetCDF library needs the 'dodsC' version, again this a pattern observed on some sources not an authoritative claim).

It takes some time to process and find these so they aren't automatically added to the results from a search.

Value

character vector of URLs

Index

cmip_available, [2](#)
cmip_available_legacy (cmip_available),
[2](#)
cmip_download, [2, 6](#)
cmip_download_config (cmip_download), [2](#)
cmip_filter_replicas, [3](#)
cmip_info (cmip_search), [4](#)
cmip_root_get (cmip_root_set), [4](#)
cmip_root_set, [4](#)
cmip_search, [4](#)
cmip_search(), [3–6](#)
cmip_simplify, [5](#)
cmip_size, [6](#)
cmip_unsimplify (cmip_simplify), [5](#)
cmip_url_to_list (cmip_search), [4](#)
cmip_urls, [6](#)