Package: predictsr (via r-universe)

October 18, 2026

Title Access the 'PREDICTS' Biodiversity Database

Version 0.2.1

Description Fetches the 'PREDICTS' database and relevant metadata from the Data Portal at the Natural History Museum, London <https://data.nhm.ac.uk>. Data were collated from over 400 existing spatial comparisons of local-scale biodiversity exposed to different intensities and types of anthropogenic pressures, from sites around the world. These data are described in Hudson et al. (2013) <doi:10.1002/ece3.2579>.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Imports digest, glue, httr2, jsonlite, logger

Suggests dplyr, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 4.1.0)

Config/testthat/parallel true

VignetteBuilder knitr

URL https://biodiversity-futures-lab.github.io/predictsr/, https://github.com/Biodiversity-Futures-Lab/predictsr

BugReports https://github.com/Biodiversity-Futures-Lab/predictsr/issues

Config/pak/sysreqs libssl-dev

Repository https://biodiversity-futures-lab.r-universe.dev

Date/Publication 2026-05-20 09:34:34 UTC

RemoteUrl https://github.com/biodiversity-futures-lab/predictsr

RemoteRef HEAD

RemoteSha 4780b0b45938daec3a792070e36300037223f093

Contents

.IsValidPredictsData 2
.ReadPredictsFileCache 2
.WritePredictsFileCache 3
GetColumnDescriptions 4
GetPredictsData 5
GetSitelevelSummaries 6
LoadPredictsData 7

Index **8**

`.IsValidPredictsData` *Check if a PREDICTS extract is valid.*

Description

A small set of basic checks to ensure that a PREDICTS extract is valid. These include checking the object is a dataframe, checking all the columns are valid, and checking that we have a nonzero row count.

Usage

```
.IsValidPredictsData(df)
```

Arguments

`df`	Dataframe, containing the PREDICTS extract.

Value

Boolean, `TRUE` if the dataframe is valid, `FALSE` if not.

`.ReadPredictsFileCache` *Read the PREDICTS file cache.*

Description

This internal helper function returns a list with 3 elements: 'valid', a boolean indicating if the cache is valid; 'data', a dataframe containing the cached PREDICTS data; and 'aux', the auxiliary data loaded from the cache.

Usage

```
.ReadPredictsFileCache(file_predicts, aux_file_predicts, requested_years)
```

Arguments

file_predicts	Character, the path to the saved PREDICTS database extract (as an RDS file).
aux_file_predicts	Character, the path to the saved PREDICTS database auxiliary metadata, saved as a JSON file.
requested_years	Numeric vector, the extract years to be saved.

Value

List, a named list of three elements: 'valid', a boolean indicating if the cache is valid; 'data', a dataframe containing the cached PREDICTS data; and 'aux', the auxiliary data loaded from the cache.

.WritePredictsFileCache

Writes the PREDICTS dataframe to disk.

Description

Given a PREDICTS database extract, loaded as an R dataframe, save it to disk and write the aux JSON file (*.aux.json), which stores metadata for the object. This includes the release years, the timestamp of when it was saved, some dimensions, and the SHA-256 hash of the dataframe (computed with 'digest').

Usage

```
.WritePredictsFileCache(df, file_predicts, aux_file_predicts, extract)
```

Arguments

df	Dataframe to be written to disk
file_predicts	Character, path to the desired file - should be an RDS file.
aux_file_predicts	Character, path to where the auxiliary file should be saved to disk.
extract	Numeric vector of release years to be saved.

Value

TRUE invisibly.

`GetColumnDescriptions` *Get a dataframe describing the columns in the PREDICTS database extract.*

Description

This function returns a dataframe containing the column descriptions for the PREDICTS database extract.

Usage

```
GetColumnDescriptions(...)
```

Arguments

...	extra arguments passed to read.csv.

Details

The PREDICTS - Predicting Responses of Ecological Diversity In Changing Terrestrial Systems - database contains a large number of columns, each corresponding to a variable describing the site or the observation. This function accesses the column descriptions for the PREDICTS database extract.

The column descriptions are provided as a dataframe, with each row corresponding to a column in the PREDICTS database extract.

There are two releases of the PREDICTS database, an initial release in 2016, and an additional release in 2022. The user chooses whether to pull summary data for the 2016 and/or 2022 release.

The data are provided under a CC NC (non-commercial) license, which means that they cannot be used for commercial purposes. The 2016 release is available under a CC BY-NC-SA 4.0 license, and the 2022 release is available under a CC NC (any) license.

Value

The column descriptions in the format as a dataframe.

Examples

```
descriptions <- GetColumnDescriptions()
```

`GetPredictsData` *Read the PREDICTS database into either a dataframe.*

Description

This returns the latest complete PREDICTS database extract as a dataframe.

Usage

```
GetPredictsData(extract = c(2016, 2022))
```

Arguments

`extract`	numeric, year/s corresponding to PREDICTS database releases to download. Options are 2016 or 2022. Defaults to `c(2016, 2022)` - the whole dataset.

Details

The data were collected as part of the PREDICTS project - Projecting Responses of Ecological Diversity In Changing Terrestrial Systems, and comprise of two releases. The first was in 2016, and the second in 2022. This function accesses the 2016 and/or 2022 release.

The database is provided as a dataframe, with each row corresponding to a site-level observation, and each column corresponding to a variable describing the site or the observation. The data are provided in a standardised format, with column names that are consistent across the database.

The data are provided under a CC NC (non-commercial) license, which means that they cannot be used for commercial purposes. The 2016 release is available under a CC BY-NC-SA 4.0 license, and the 2022 release is available under a CC NC (any) license.

Value

A dataframe containing the v1.1 PREDICTS database extract/s.

Examples

```
predicts <- GetPredictsData()
predicts_2016 <- GetPredictsData(extract = 2016)
```

`GetSitelevelSummaries` *Get the PREDICTS database site level summaries.*

Description

This acesses summary data for the relevant PREDICTS database extract.

Usage

```
GetSitelevelSummaries(extract = c(2016, 2022))
```

Arguments

`extract`	Numeric, year/s corresponding to PREDICTS database releases to download. Options are 2016 or 2022. Defaults to `c(2016, 2022)` - the whole dataset.

Details

The PREDICTS database contains site-level summaries of the data collected as part of the PREDICTS project - Projecting Responses of Ecological Diversity In Changing Terrestrial Systems.

The site-level summaries are provided as a dataframe, with each row corresponding to a site-level observation, and each column corresponding to a variable describing the site or the observation. The data are provided in a standardised format, with column names that are consistent across the database.

There are two releases of the PREDICTS database, an initial release in 2016, and an additional release in 2022. The user chooses whether to pull summary data for the 2016 and/or 2022 release.

The data are provided under a CC NC (non-commercial) license, which means that they cannot be used for commercial purposes. The 2016 release is available under a CC BY-NC-SA 4.0 license, and the 2022 release is available under a CC NC (any) license.

Value

The site-level summary data as a dataframe.

Examples

```
summaries <- GetSitelevelSummaries()
summaries_2016 <- GetSitelevelSummaries(extract = 2016)
```

`LoadPredictsData` *Load (or download) PREDICTS data to a user-specified RDS file.*

Description

Implements a simple file-based cache. You supply a target filename (e.g. "data/predicts_2016_2022.rds"). The function will:

1. Look for that RDS file and the companion metadata file "filename.aux.json" (e.g. "data/predicts_2016_2022.rds.aux.jso
2. If both exist, verify the file hash, minimal structure, and requested years.
3. If validation passes return the loaded object.
4. Otherwise download fresh data via `GetPredictsData(extract)`, overwrite the RDS, write a new `.aux.json`, and return the dataframe.

The data are provided under a CC NC (non-commercial) license, which means that they cannot be used for commercial purposes. The 2016 release is available under a CC BY-NC-SA 4.0 license, and the 2022 release is available under a CC NC (any) license.

Usage

```
LoadPredictsData(file_predicts, extract = c(2016, 2022), force_refresh = FALSE)
```

Arguments

`file_predicts`	Character path to the desired PREDICTS database RDS file (must end with ".rds").
`extract`	Integer vector of release years to fetch. Defaults to `c(2016, 2022)`.
`force_refresh`	Logical; if TRUE always re-download and overwrite existing files.

Value

A dataframe containing the requested PREDICTS extract.

Examples

```
file_predicts <- file.path(tempdir(), "predicts.rds")
df_predicts <- LoadPredictsData(file.path(tempdir(), "predicts.rds"))
```

Index

.IsValidPredictsData, 2
.ReadPredictsFileCache, 2
.WritePredictsFileCache, 3

GetColumnDescriptions, 4
GetPredictsData, 5
GetSitelevelSummaries, 6

LoadPredictsData, 7