

Package ‘rdefra’

July 8, 2018

Title Interact with the UK AIR Pollution Database from DEFRA

Version 0.3.5

Maintainer Claudia Vitolo <cvitolodev@gmail.com>

URL <https://github.com/ropensci/rdefra>

BugReports <https://github.com/ropensci/rdefra/issues>

Description Get data from DEFRA's UK-AIR website <<https://uk-air.defra.gov.uk/>>. It basically scrapes the HTML content.

Depends R (>= 2.10)

Imports lubridate, tibble, httr, xml2, dplyr, sp, rgdal

Suggests testthat, leaflet, parallel, rmarkdown, knitr, ggplot2, zoo, lintr

VignetteBuilder knitr

LazyData true

Encoding UTF-8

License GPL-3

Repository CRAN

RoxygenNote 6.0.1

NeedsCompilation no

Author Claudia Vitolo [aut, cre],

Andrew Russell [ctb] (This package is part of the KEHRA project and Andrew supervised the technical/software development.),

Allan Tucker [ctb] (This package is part of the KEHRA project and Allan supervised the technical/software development.),

Maëlle Salmon [ctb] (Maëlle Salmon reviewed the package for rOpenSci, see <https://github.com/ropensci/onboarding/issues/68>),

Hao Zhu [ctb] (Hao Zhu reviewed the package for rOpenSci, see <https://github.com/ropensci/onboarding/issues/68>)

Date/Publication 2018-07-08 14:30:03 UTC

R topics documented:

rdefra	2
stations	2
ukair_catalogue	3
ukair_get_coords	5
ukair_get_hourly_data	6
ukair_get_site_id	7

Index	8
--------------	----------

rdefra	<i>rdefra: Interact with the UK AIR Pollution Database from DEFRA</i>
--------	---

Description

The R package `rdefra` allows to retrieve air pollution data from the Air Information Resource (UK-AIR) of the Department for Environment, Food and Rural Affairs in the United Kingdom (see <https://uk-air.defra.gov.uk/>). UK-AIR does not provide public APIs for programmatic access to data, therefore this package scrapes the HTML pages to get relevant information.

stations	<i>List of all the DEFRA air quality monitoring stations with complete coordinates</i>
----------	--

Description

This is the list of all the air quality monitoring stations ever installed in the UK and operated by DEFRA networks (as per February 2016). As the network expands, metadata for new stations will be added.

Usage

```
data("stations")
```

Format

A data frame with 6561 observations on the following 14 variables.

UK.AIR.ID ID reference for monitoring stations

EU.Site.ID EU.Site.ID

EMEP.Site.ID EMEP.Site.ID

Site.Name Site name

Environment.Type a factor with levels Background Rural Background Suburban Background Urban Industrial Suburban Industrial Unknown Industrial Urban Traffic Urban Unknown Unknown

Zone Zone
 Start.Date Start date
 End.Date End date
 Latitude Latitude (WGS 84)
 Longitude Longitude (WGS 84)
 Altitude..m. Altitude in metres above sea level
 Networks Monitoring Networks
 AURN.Pollutants.Measured Pollutant measured
 Site.Description Site.Description
 Easting Easting coordinate (British National Grid)
 Northing Northing coordinate (British National Grid)
 SiteID Site ID

Source

<http://uk-air.defra.gov.uk/>

ukair_catalogue

Get DEFRA UK-AIR stations metadata

Description

This function fetches the catalogue of monitoring stations from DEFRA's website.

Usage

```
ukair_catalogue(site_name = "", pollutant = 9999, group_id = 9999,
  closed = "true", country_id = 9999, region_id = 9999)
```

Arguments

site_name	This is the name of a specific site. By default this is left blank to get info on all the available sites.
pollutant	This is a number from 1 to 10. Default is 9999, which means all the pollutants.
group_id	This is the identification number of a group of stations. Default is 9999 which means all available networks.
closed	This is "true" to include closed stations, "false" otherwise.
country_id	This is the identification number of the country, it can be a number from 1 to 6. Default is 9999, which means all the countries.
region_id	This is the identification number of the region. 1 = Aberdeen City, etc. (for the full list see https://uk-air.defra.gov.uk/). Default is 9999, which means all the local authorities.

Details

The argument `Pollutant` is defined based on the following convention:

- 1 = Ozone (O₃)
- 2 = Nitrogen oxides (NO_x)
- 3 = Carbon monoxide (CO)
- 4 = Sulphur dioxide (SO₂)
- 5 = Particulate Matter (PM₁₀)
- 6 = Particulate Matter (PM_{2.5})
- 7 = PAHs
- 8 = Metals in PM₁₀
- 9 = Benzene
- 10 = Black Carbon

The argument `group_id` is defined based on the following convention:

- 1 = UKEAP: Precip-Net
- 2 = Air Quality Strategy Pollutants
- 3 = Ammonia and Nitric Acid
- 4 = Automatic Urban and Rural Monitoring Network (AURN)
- 5 = Dioxins and Furans
- 6 = Black Smoke & SO₂
- 7 = Automatic Hydrocarbon Network
- 8 = Heavy Metals
- 9 = Nitrogen Dioxide Diffusion Tube
- 10 = PAH Andersen
- 11 = Particle Size Composition
- 12 = PCBs
- 13 = TOMPs
- 14 = Non-Automatic Hydrocarbon Network
- 15 = 1,3-Butadiene Diffusion Tube
- 16 = Black Carbon
- 17 = Automatic Urban and Rural Monitoring Network (AURN)
- 18 = Defra NO₂ Diffusion Tube
- 19 = PAH Digitel (solid phase)
- 20 = PAH Digitel (solid+vapour)
- 21 = PAH Deposition
- 22 = Particle size and number
- 23 = Rural Automatic Mercury network

- 24 = Urban Sulphate
- 25 = UKEAP: Rural NO2
- 26 = Automatic Urban and Rural Monitoring Network (AURN)
- 27 = UKEAP: National Ammonia Monitoring Network
- 28 = UKEAP: Acid Gases & Aerosol Network
- 29 = Particle Speciation (MARGA)
- 30 = UKEAP: Historic Aerosol measurements

The argument `country_id` is defined based on the following convention:

- 1 = England
- 2 = Wales
- 3 = Scotland
- 4 = Northern Ireland
- 5 = Republic of Ireland
- 6 = Channel Islands

Value

A named vector containing Easting and Northing coordinates.

Examples

```
## Not run:  
stations <- ukair_catalogue()  
  
## End(Not run)
```

<code>ukair_get_coords</code>	<i>Get Easting and Northing coordinates from DEFRA</i>
-------------------------------	--

Description

This function takes as input the UK AIR ID and returns Easting and Northing coordinates (British National Grid, EPSG:27700).

Usage

```
ukair_get_coords(ids)
```

Arguments

<code>ids</code>	contains the station identification code defined by DEFRA. It can be: a) an alphanumeric string, b) a vector of strings or c) a data frame. In the latter case, the column containing the codes should be named "UK.AIR.ID", all the other columns will be ignored.
------------------	---

Details

If the input is a data frame with some of the columns named "UK.AIR.ID", "Latitude" and "Longitude", the function only infills missing Latitude/Longitude values.

Value

A data.frame containing at least five columns named "UK.AIR.ID", "Easting", "Northing", "Latitude" and "Longitude".

Examples

```
## Not run:
# Case a: alphanumeric string
ukair_get_coords("UKA12536")

# Case b: vector of strings
ukair_get_coords(c("UKA15910", "UKA15956", "UKA16663", "UKA16097"))

# Case c: data frame
ukair_get_coords(ukair_catalogue()[1:10,])

## End(Not run)
```

ukair_get_hourly_data *Get hourly data for DEFRA stations*

Description

This function fetches hourly data from DEFRA's air pollution monitoring stations.

Usage

```
ukair_get_hourly_data(site_id = NULL, years = NULL)
```

Arguments

site_id	This is the ID of a specific site.
years	Years for which data should be downloaded.

Details

The measurements are generally in $\mu\text{g}/\text{m}^3$ (micrograms per cubic metre). To check the units, refer to the table of attributes (see example below). Please double check the units on the DEFRA website, as they might change over time.

Value

A data.frame containing hourly pollution data.

Examples

```
## Not run:
# Get data for 1 year
output <- ukair_get_hourly_data("ABD", 2014)

# Get data for multiple years
output <- ukair_get_hourly_data("ABD", 2014:2016)

# Get units
attributes(output)$units

## End(Not run)
```

ukair_get_site_id *Get site identification numbers for DEFRA stations*

Description

Given the UK AIR ID (from the `ukair_catalogue()`), this function fetches the catalogue of monitoring stations from DEFRA's website.

Usage

```
ukair_get_site_id(id_s)
```

Arguments

`id_s` An alphanumeric string (or vector of strings) containing the UK AIR ID defined by DEFRA.

Value

A named vector containing the site `id_s`.

Examples

```
## Not run:
ukair_get_site_id("UKA00399")

## End(Not run)
```

Index

*Topic **datasets**

stations, [2](#)

rdefra, [2](#)

rdefra-package (rdefra), [2](#)

stations, [2](#)

ukair_catalogue, [3](#)

ukair_get_coords, [5](#)

ukair_get_hourly_data, [6](#)

ukair_get_site_id, [7](#)