Package ‘rdefra’

April 6, 2020

Title Interact with the UK AIR Pollution Database from DEFRA
Version 0.3.8
Maintainer Claudia Vitolo <cvitolodev@gmail.com>
URL https://docs.ropensci.org/rdefra,
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, rmarkdown, knitr, ggplot2, zoo
SystemRequirements GDAL
VignetteBuilder knitr
LazyData true
Encoding UTF-8
License GPL-3
Repository CRAN
RoxygenNote 7.1.0
NeedsCompilation no
Author Claudia Vitolo [aut, cre] (<https://orcid.org/0000-0002-4252-1176>),
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 2020-04-06 09:50:02 UTC
R topics documented:

rdefra: Interact with the UK AIR Pollution Database from DEFRA

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

List of all the DEFRA air quality monitoring stations with complete coordinates

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
**ukair_catalogue**

---

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone</td>
<td></td>
</tr>
<tr>
<td>Start.Date</td>
<td>Start date</td>
</tr>
<tr>
<td>End.Date</td>
<td>End date</td>
</tr>
<tr>
<td>Latitude</td>
<td>Latitude (WGS 84)</td>
</tr>
<tr>
<td>Longitude</td>
<td>Longitude (WGS 84)</td>
</tr>
<tr>
<td>Northing</td>
<td>Northing coordinate (British National Grid)</td>
</tr>
<tr>
<td>Easting</td>
<td>Easting coordinate (British National Grid)</td>
</tr>
<tr>
<td>Altitude</td>
<td>Altitude in metres above sea level</td>
</tr>
<tr>
<td>Networks</td>
<td>Monitoring Networks</td>
</tr>
<tr>
<td>AURN.Pollutants.Measured</td>
<td>Pollutant measured</td>
</tr>
<tr>
<td>Site.Description</td>
<td>Description of the site.</td>
</tr>
<tr>
<td>SiteID</td>
<td>Site ID, used to retrieve time series data.</td>
</tr>
</tbody>
</table>

**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**

```r
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 (NOx)
- 3 = Carbon monoxide (CO)
- 4 = Sulphur dioxide (SO2)
- 5 = Particulate Matter (PM10)
- 6 = Particulate Matter (PM2.5)
- 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 Digital (solid phase)
• 20 = PAH Digital (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 dataframe listing stations and related information.

Examples

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

## End(Not run)
```

---

**ukair_get_coordinates**  
Get Easting and Northing coordinates from DEFRA

**Description**

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

ukair_get_coordinates(ids)

Arguments

ids 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", "Northing" and "Easting", the function only infills missing Northing/Easting values (if available on the relevant webpage).

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_coordinates("UKA12536")

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

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

## End(Not run)
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 \( \mu g/m^3 \) (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

```r
## 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

```r
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

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

## End(Not run)
```
Index

*Topic datasets
  stations, 2
  rdefra, 2
  stations, 2

  ukair_catalogue, 3
  ukair_get_coordinates, 5
  ukair_get_hourly_data, 6
  ukair_get_site_id, 7