

Package ‘phsopendata’

September 1, 2025

Title Extract from the Scottish Health and Social Care Open Data Platform

Version 1.0.0

Description Extract and interact with data from the Scottish Health and Social Care Open Data platform <<https://www.opendata.nhs.scot>>.

License MIT + file LICENSE

URL <https://github.com/Public-Health-Scotland/phsopendata>,
<https://public-health-scotland.github.io/phsopendata/>

BugReports <https://github.com/Public-Health-Scotland/phsopendata/issues>

Imports cli (>= 3.2.0), dplyr (>= 1.0.0), httr (>= 1.0.0), magrittr (>= 1.0.0), purrr (>= 1.0.0), rlang (>= 1.0.0), stringdist, tibble (>= 3.0.0)

Suggests covr, jsonlite (>= 1.1), readr (>= 1.0.0), testthat (>= 3.0.0), xml2

Config/testthat/edition 3

Config/testthat/parallel true

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Public Health Scotland [cph],
Csilla Scharle [aut, cre],
James Hayes [aut] (ORCID: <<https://orcid.org/0000-0002-5380-2029>>),
David Aikman [aut],
Ross Hull [aut]

Maintainer Csilla Scharle <csilla.scharle2@phs.scot>

Repository CRAN

Date/Publication 2025-09-01 09:50:02 UTC

Contents

get_dataset	2
get_dataset_additional_info	3
get_latest_resource	4
get_resource	5
get_resource_sql	6
list_datasets	7
list_resources	8
Index	9

get_dataset	<i>Get Open Data resources from a dataset</i>
-------------	---

Description

Downloads multiple resources from a dataset on the NHS Open Data platform by dataset name, with optional row limits and context columns.

Usage

```
get_dataset(
  dataset_name,
  max_resources = NULL,
  rows = NULL,
  row_filters = NULL,
  col_select = NULL,
  include_context = FALSE
)
```

Arguments

dataset_name	Name of the dataset as found on NHS Open Data platform (character).
max_resources	(optional) The maximum number of resources to return (integer). If not set, all resources are returned.
rows	(optional) Maximum number of rows to return (integer).
row_filters	(optional) A named list or vector specifying values of columns/fields to keep (e.g., list(Date = 20220216, Sex = "Female")).
col_select	(optional) A character vector containing the names of desired columns/fields (e.g., c("Date", "Sex")).
include_context	(optional) If TRUE, additional information about the resource will be added as columns to the data, including the resource ID, the resource name, the creation date, and the last modified/updated date.

Value

A [tibble](#) with the data.

See Also

[get_resource\(\)](#) for downloading a single resource from a dataset.

Examples

```
get_dataset("gp-practice-populations", max_resources = 2, rows = 10)
```

`get_dataset_additional_info`
get a datasets additional info

Description

`get_dataset_additional_info()` returns a tibble of dataset names along with the amount of resources it has and the date it was last updated. Last updated is taken to mean the most recent date a resource within the dataset was created or modified.

Usage

```
get_dataset_additional_info(dataset_name)
```

Arguments

`dataset_name` Name of the dataset as found on [NHS Open Data platform](#) (character).

Value

a [tibble](#) with the data

Examples

```
get_dataset_additional_info("gp-practice-populations")
```

get_latest_resource *Get the latest resource from a data set*

Description

Returns the latest resource available in a dataset.

Usage

```
get_latest_resource(  
  dataset_name,  
  rows = NULL,  
  row_filters = NULL,  
  col_select = NULL,  
  include_context = TRUE  
)
```

Arguments

dataset_name	Name of the dataset as found on NHS Open Data platform (character).
rows	(optional) Maximum number of rows to return (integer).
row_filters	(optional) A named list or vector specifying values of columns/fields to keep (e.g., list(Date = 20220216, Sex = "Female")).
col_select	(optional) A character vector containing the names of desired columns/fields (e.g., c("Date", "Sex")).
include_context	(optional) If TRUE, additional information about the resource will be added as columns to the data, including the resource ID, the resource name, the creation date, and the last modified/updated date.

Details

There are some datasets on the open data platform that keep historic resources instead of updating existing ones. For these it is useful to be able to retrieve the latest resource. As of 1.8.2024 these data sets include:

- gp-practice-populations
- gp-practice-contact-details-and-list-sizes
- nhsscotland-payments-to-general-practice
- dental-practices-and-patient-registrations
- general-practitioner-contact-details
- prescribed-dispensed
- dispenser-location-contact-details
- community-pharmacy-contractor-activity

Value

a [tibble](#) with the data

Examples

```
dataset_name <- "gp-practice-contact-details-and-list-sizes"

data <- get_latest_resource(dataset_name)

filters <- list("Postcode" = "DD11 1ES")
wanted_cols <- c("PracticeCode", "Postcode", "Dispensing")

filtered_data <- get_latest_resource(
  dataset_name = dataset_name,
  row_filters = filters,
  col_select = wanted_cols
)
```

get_resource *Get Open Data resource*

Description

Downloads a single resource from the NHS Open Data platform by resource ID, with optional filtering and column selection.

Usage

```
get_resource(
  res_id,
  rows = NULL,
  row_filters = NULL,
  col_select = NULL,
  include_context = FALSE
)
```

Arguments

res_id	The resource ID as found on NHS Open Data platform (character).
rows	(optional) Maximum number of rows to return (integer).
row_filters	(optional) A named list or vector specifying values of columns/fields to keep (e.g., list(Date = 20220216, Sex = "Female")).
col_select	(optional) A character vector containing the names of desired columns/fields (e.g., c("Date", "Sex")).
include_context	(optional) If TRUE, additional information about the resource will be added as columns to the data, including the resource ID, the resource name, the creation date, and the last modified/updated date.

Value

A [tibble](#) with the data.

See Also

[get_dataset\(\)](#) for downloading all resources from a given dataset.

Examples

```
res_id <- "ca3f8e44-9a84-43d6-819c-a880b23bd278"

data <- get_resource(res_id)

filters <- list("HB" = "S08000030", "Month" = "202109")
wanted_cols <- c("HB", "Month", "TotalPatientsSeen")

filtered_data <- get_resource(
  res_id = res_id,
  row_filters = filters,
  col_select = wanted_cols
)
```

get_resource_sql

Get PHS Open Data using SQL

Description

Downloads data from the NHS Open Data platform using a SQL query. Similar to [get_resource\(\)](#), but allows more flexible server-side querying. This function has a lower maximum row number (32,000 vs 99,999) for returned results.

Usage

```
get_resource_sql(sql)
```

Arguments

`sql` A single PostgreSQL SELECT query (character). Must include a resource ID, which must be double-quoted (e.g., `SELECT * from "58527343-a930-4058-bf9e-3c6e5cb04010"`).

Value

A [tibble](#) with the query results. Only 32,000 rows can be returned from a single SQL query.

See Also

[get_resource\(\)](#) for downloading a resource without using a SQL query.

Examples

```
sql <- "
  SELECT
    \"TotalCancelled\", \"TotalOperations\", \"Hospital\", \"Month\"
  FROM
    \"bcc860a4-49f4-4232-a76b-f559cf6eb885\"
  WHERE
    \"Hospital\" = 'D102H'
"
df <- get_resource_sql(sql)

# This is equivalent to:
cols <- c("TotalCancelled", "TotalOperations", "Hospital", "Month")
row_filter <- c(Hospital = "D102H")

df2 <- get_resource(
  "bcc860a4-49f4-4232-a76b-f559cf6eb885",
  col_select = cols,
  row_filters = row_filter
)
```

list_datasets

Lists all available datasets

Description

list_datasets() shows all of the datasets hosted on the phs open data platform.

Usage

```
list_datasets()
```

Value

A tibble.

Examples

```
head(list_datasets())
```

list_resources	<i>Lists all available resources for a dataset</i>
----------------	--

Description

list_resources() returns all of the resources associated with a dataset

Usage

```
list_resources(dataset_name)
```

Arguments

dataset_name Name of the dataset as found on [NHS Open Data platform](#) (character).

Value

a [tibble](#) with the data

Examples

```
list_resources("weekly-accident-and-emergency-activity-and-waiting-times")
```


Index

`get_dataset`, 2
`get_dataset()`, 6
`get_dataset_additional_info`, 3
`get_latest_resource`, 4
`get_resource`, 5
`get_resource()`, 3, 6
`get_resource_sql`, 6

`list_datasets`, 7
`list_resources`, 8

`tibble`, 3, 5, 6, 8