

Package ‘phsopendata’

May 6, 2026

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

Version 1.1.0

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), lifecycle, magrittr (>= 1.0.0), purrr (>= 1.0.0), readr (>= 1.0.0), rlang (>= 1.0.0), stringdist, tibble (>= 3.0.0)

Suggests covr, curl, jsonlite (>= 1.1), spelling, testthat (>= 3.0.0), xml2

Config/roxygen2/markdown TRUE

Config/roxygen2/version 8.0.0

Config/testthat/edition 3

Encoding UTF-8

Language en-GB

NeedsCompilation no

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

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

Repository CRAN

Date/Publication 2026-05-06 08:50:02 UTC

Contents

| | |
|-----------------------------|---|
| get_dataset | 2 |
| get_dataset_additional_info | 3 |
| get_latest_resource | 4 |
| get_resource | 5 |
| get_resource_sql | 6 |
| list_datasets | 8 |
| list_resources | 8 |

| | |
|--------------|-----------|
| Index | 10 |
|--------------|-----------|

| | |
|-------------|-----------------------------------------------|
| 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 the 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

```
## Not run:  
get_dataset("gp-practice-populations", max_resources = 2, rows = 10)  
  
## End(Not run)
```

`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 the [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 the 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

```
## Not run:
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
)

## End(Not run)
```

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 SELECT query (character).

Details

Only 32,000 rows can be returned from a single SQL query.

Value

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

SQL syntax

The resource ID must be double-quoted, e.g. `SELECT * FROM "58527343-a930-4058-bf9e-3c6e5cb04010"`, as must column names, e.g. `"Year"`. Strings require single quotes, e.g. `'value'`. This syntax is needed because the CKAN DataStore uses [PostgreSQL](#).

R raw strings

Enclosing the query in an R raw string avoids the need to escape embedded quotes, e.g. `\\"TotalCancelled\\"`. Square brackets are the recommended delimiter, i.e. `r"[...]"`, because `)` within a query, e.g. `SUM("TotalCancelled")`, would prematurely close the string. Another option is `r"{...}"`, in the rare case that the query contains `]`.

See Also

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

Examples

```
# Basic query
cancelled_ops <- get_resource_sql(r"[
SELECT
  "TotalCancelled",
  "TotalOperations",
  "Hospital",
  "Month"
FROM
  "bcc860a4-49f4-4232-a76b-f559cf6eb885"
WHERE
  "Hospital" = 'D102H'
]")

# Joining two resources
hb_pop <- get_resource_sql(r"[
SELECT
  pops."Year",
  pops."HB",
  lookup."HBName",
  pops."AllAges"
FROM
  "27a72cc8-d6d8-430c-8b4f-3109a9ceadb1" AS pops
  JOIN "652ff726-e676-4a20-abda-435b98dd7bdc" AS lookup ON pops."HB" = lookup."HB"
WHERE
  pops."Sex" = 'All'
  AND pops."Year" > 2006
]")
```

| | |
|---------------|-------------------------------------|
| list_datasets | <i>Lists all available datasets</i> |
|---------------|-------------------------------------|

Description

list_datasets() shows all of the datasets hosted on the PHS Open Data Platform.

Usage

```
list_datasets()
```

Value

A tibble.

Lifecycle

[Superseded]

list_datasets() has been superseded by list_resources(). While list_datasets() only returns a list of dataset packages, list_resources() provides a more comprehensive and flexible interface for exploring the PHS Open Data platform. It returns both datasets and their associated resources in a single tibble, and supports filtering by dataset titles or resource names.

See Also

[list_resources\(\)](#)

Examples

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

| | |
|----------------|-----------------------------------------------------------------------------------|
| list_resources | <i>Provides an overview of all resources available on the Open Data platform.</i> |
|----------------|-----------------------------------------------------------------------------------|

Description

Provides an overview of all resources available from opendata.nhs.scot, with the option to limit results based on both dataset and resource names. The returned tibble can be used to look-up dataset and resource ids, and is useful for exploring the available data sets.

Usage

```
list_resources(  
  dataset_contains = NULL,  
  resource_contains = NULL,  
  dataset_name = lifecycle::deprecated()  
)
```

Arguments

`dataset_contains` A character string containing an expression to be used as search criteria against the dataset name.

`resource_contains` A character string containing a **regular expression** to be matched against available resource names.

`dataset_name` Deprecated. Use `dataset_contains` instead.

Value

A [tibble](#) containing details of all available datasets and resources, or those containing the string specified in the `dataset_contains` and `resource_contains` arguments.

Examples

```
list_resources()  
list_resources(dataset_contains = "standard-populations")  
list_resources(  
  dataset_contains = "standard-populations", resource_contains = "European"  
)
```

Index

`get_dataset`, 2
`get_dataset()`, 6
`get_dataset_additional_info`, 3
`get_latest_resource`, 4
`get_resource`, 5
`get_resource()`, 3, 6, 7
`get_resource_sql`, 6

`list_datasets`, 8
`list_resources`, 8
`list_resources()`, 8

`tibble`, 3, 5–7, 9