

Package ‘sooty’

May 22, 2025

Title Data Source Catalogues Online for Southern Ocean Ecosystem Research

Version 0.5.0

Description Obtains lists of files of remote sensing collections for Southern Ocean surface properties. Commonly used data sources of sea surface temperature, sea ice concentration, and altimetry products such as sea surface height and sea surface currents are cached in object storage on the Pawsey Supercomputing Research Centre facility. Patterns of working to retrieve data from these object storage catalogues are described. The catalogues include complete collections of datasets Reynolds et al. (2008) ``NOAA Optimum Interpolation Sea Surface Temperature (OISST) Analysis, Version 2.1" <[doi:10.7289/V5SQ8XB5](https://doi.org/10.7289/V5SQ8XB5)>, Spreen et al. (2008) ``Artist Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) sea ice concentration" <[doi:10.1029/2005JC003384](https://doi.org/10.1029/2005JC003384)>. In future releases helpers will be added to identify particular data collections and target specific dates for earth observation data for reading, as well as helpers to retrieve data set citation and provenance details. This work was supported by resources provided by the Pawsey Supercomputing Research Centre with funding from the Australian Government and the Government of Western Australia. This software was developed by the Integrated Digital East Antarctica program of the Australian Antarctic Division.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

RoxygenNote 7.3.2

Imports arrow, dplyr, S7, tibble

URL <https://github.com/mdsumner/sooty>

BugReports <https://github.com/mdsumner/sooty/issues>

Suggests curl, spelling, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Michael D. Sumner [aut, cre],
Aleks Terauds [cph, ctb] (Provided logo photo from p116 of
'Subantarctic wilderness: Macquarie Island, 2007(978-1741753028)')

Maintainer Michael D. Sumner <michael.sumner@aad.gov.au>

Repository CRAN

Date/Publication 2025-05-22 09:30:02 UTC

Contents

| | |
|------------------------------|----------|
| available_datasets | 2 |
| datasource | 3 |
| sooty_files | 4 |
| Index | 5 |

| | |
|--------------------|--------------------------------|
| available_datasets | <i>List available datasets</i> |
|--------------------|--------------------------------|

Description

In sooty_files() the data source files are grouped by Dataset, this is the list of unique datasets, values that can be used in datasource(<name>).

Usage

available_datasets()

Value

character vector of available dataset ids for datasource()

Examples

available_datasets()

| | |
|------------|---|
| datasource | <i>Create a datasource object. A data source provides a list of files that together comprise a dataset.</i> |
|------------|---|

Description

Generates an object whose "@id" property may be set, which then communicates with a dataset of files/objects that sooty knows about.

Usage

```
datasource(id = NA_character_)
```

```
dataset(...)
```

Arguments

| | |
|-----|---|
| id | a dataset label, see 'datasource()@available_datasources' (get, and settable) |
| ... | only used by deprecated function, will become defunct |

Details

Compare 'curated' to 'sooty_files(curated = FALSE)', if it is curated sooty knows what dataset it belongs to, and otherwise it's just the huge list of files we're interested in for our work. All of the curation is done outside of sooty.

The following properties are available via the @ slot:

- n the number of files (objects) comprising the dataset (get, not settable)
- mindate the minimum available date for the files
- 'maxdate1 the maximum available date for the files
- 'source1 the set of files (objects) belonging to this dataset (get, not settable)

Note

This was originally called dataset() which usage has now been deprecated.

Examples

```
## available dataset names
if (interactive()) {
  available_datasets()
}
## set to one of those
ds <- datasource("ghrsst-tif")
## access the 'ds@source' slot, files with 'date', 'source' (GDAL-readable)
```

| | |
|-------------|--|
| sooty_files | <i>Obtain object storage catalogues as a dataframe of file/object identifiers.</i> |
|-------------|--|

Description

The object (file) catalogue of available sources is stored in Parquet format on Pawsey object storage. This function retrieves the curated catalogue, or the raw catalogue.

Usage

```
sooty_files(curated = TRUE)
```

Arguments

| | |
|---------|--|
| curated | logical TRUE by default, set to FALSE to return raw object catalogue |
|---------|--|

Details

In the curated case, the returned data frame has columns 'date', 'source' which are the main useful fields, these describe the date of the data in the file, and its full URI (Uniform Resource Identifier) source on S3 object storage. There are also fields 'Bucket', 'Key', and 'protocol' from which 'source' is constructed.

The original publisher URI can be reconstructed by replacing the value of 'protocol' in 'source' with 'https://'.

The public object URI can be reconstructed by replacing the value of 'protocol' in 'source' with 'https://projects.pawsey.org.au'.

Value

a data frame, see details

Examples

```
if (interactive()) {  
  sooty_files(FALSE)  
}  
  
sooty_files()
```

Index

`available_datasets`, [2](#)

`dataset (datasource)`, [3](#)
`datasource`, [3](#)

`sooty_files`, [4](#)