Package 'sooty'

March 10, 2026

Title Data Source Catalogues Online for Southern Ocean Ecosystem
Research

Version 0.6.1

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 re-
trieve data from these object storage
catalogues are described. The catalogues include complete collec-
tions of datasets Reynolds et al. (2008)
``NOAA Optimum Interpolation Sea Surface Temperature (OISST) Analysis, Ver-
sion 2.1'' <doi:10.7289/V5SQ8XB5>,
Spreen et al. (2008) ``Artist Advanced Microwave Scanning Radiometer for Earth Observ-
ing System (AMSR-E)
sea ice concentration'' <doi:10.1029/2005JC003384>. In future re-
leases 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 re-
sources provided
by the Pawsey Supercomputing Research Centre with funding from the Australian Govern-
ment and the Government of
Western Australia. This software was developed by the Integrated Digital East Antarctica pro-
gram of the Australian
Antarctic Division.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

Imports arrow, curl, S7, tibble

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

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

Suggests spelling, testthat (>= 3.0.0), withr

Config/testthat/edition 3

Depends R (>= 2.10)

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 2026-03-10 06:10:08 UTC

Contents

---

   available_datasets   *List available datasets*

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(id).

Usage

```
available_datasets()
```

Value

character vector of available dataset ids for datasource()

Examples

```
op <- options("sooty.allow.cache" = FALSE)
available_datasets()
options(op)
```

datasource	*Create a datasource object. A data source provides a list of files that together comprise a dataset.*

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 [available_datasets()](#)
...	only used by deprecated function, will become defunct

Details

The following properties are available via the @ slot:

- id a dataset label, see [available_datasets()](#) (gettable and settable)
- n the number of files (objects) comprising the dataset (get only)
- mindate the minimum available date for the files (get only)
- maxdate the maximum available date for the files (get only)
- source the set of files (objects) belonging to this dataset (get only)

By default sooty maintains a local cache of the catalogue used to populate these properties. Set options("sooty.allow.cache" = FALSE) to use only the bundled sysdata, or options("sooty.cache.path" = tempdir()) to redirect the cache directory. See [sooty_cache_info()](#) for details.

Note

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

Examples

```
op <- options("sooty.allow.cache" = FALSE)
## available dataset names
available_datasets()
## set to one of those
ds <- datasource("ghrsst-tif")
options(op)
```

---

sooty_cache_info *Show sooty cache status*

Description

Reports the active cache configuration, including the effect of any options that have been set. See the Options section below for details.

Usage

```
sooty_cache_info()
```

Value

A data frame (invisibly) with cache details.

Options

Two options control cache behaviour:

sooty.allow.cache logical, default TRUE. Set to FALSE to skip all disk I/O and use only the bundled sysdata. Suitable for examples, tests, and offline use: `options("sooty.allow.cache" = FALSE)`.

sooty.cache.path path, default `tools::R_user_dir("sooty", "cache")`. Override the cache directory. Useful for CI or shared environments: `options("sooty.cache.path" = tempdir())`.

Examples

```
sooty_cache_info()

op <- options("sooty.allow.cache" = FALSE)
sooty_cache_info()
options(op)
```

---

sooty_files *Obtain object storage catalogues as a dataframe of file/object identifiers.*

Description

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

Usage

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

Arguments

curated   logical TRUE by default, ignored with a warning if FALSE

Details

The returned curated 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'.

By default sooty maintains a local cache of the catalogue, refreshed once per session when internet is available. Set options("sooty.allow.cache" = FALSE) to suppress all disk I/O and use only the bundled sysdata, or options("sooty.cache.path" = tempdir()) to redirect the cache directory. See sooty_cache_info() for details.

Value

a data frame, see details

Examples

```
op <- options("sooty.allow.cache" = FALSE)
sooty_files()
options(op)
```

Index