Package 'SpatialData.data'

November 25, 2026

Title SpatialData datasets

Depends R (>= 4.6.0)

Version 1.0.0

Description A companion package to access scverse-curated example SpatialData stores. The datasets span spatial omics technologies such as Visium, Visium HD, Xenium, MERFISH, MIBI-TOF, and imaging mass cytometry, and are distributed as zipped Zarr stores through the Bioconductor bucket on the NSF Open Storage Network (OSN) or the scverse spatialdata sandbox. Available datasets and their metadata can be listed, and each dataset is downloaded once, cached locally with 'BiocFileCache', and read into R as a SpatialData object via 'spatialdataR' package.

Imports utils, BiocFileCache, spatialdataR, paws.storage

Suggests BiocStyle, knitr, testthat

biocViews DataImport, DataRepresentation, Infrastructure, ImmunoOncology, GeneExpression, Transcriptomics, SingleCell, Spatial

License Artistic-2.0

Encoding UTF-8

VignetteBuilder knitr

URL https://github.com/HelenaLC/SpatialData.data, https://helenalc.github.io/SpatialData.data/

BugReports https://github.com/HelenaLC/SpatialData.data/issues

Config/roxygen2/version 8.1.0

git_url https://git.bioconductor.org/packages/SpatialData.data

git_branch RELEASE_3_24

git_last_commit ba6d4c8

git_last_commit_date 2026-11-23

Repository Bioconductor 3.24

Date/Publication 2026-11-24

Author Helena L. Crowell [aut, cre] (ORCID: <https://orcid.org/0000-0002-4801-1767>), Artür Manukyan [aut] (ORCID: <https://orcid.org/0000-0002-0441-9517>), Hugo Gruson [aut] (ORCID: <https://orcid.org/0000-0002-4094-1476>), Vince Carey [aut] (ORCID: <https://orcid.org/0000-0003-4046-0063>)

Maintainer Helena L. Crowell <helena@crowell.eu>

Contents

SD.data_available 2
SD.data_list 3
SD.data_load 3

Index **5**

`SD.data_available` *SD.data_available*

Description

Function for interrogating files across buckets. Please use paws.storage::s3' to interrogate buckets for zipped zarr archives or raw readouts for various platforms.

Usage

```
SD.data_available(src = "sandbox")
```

Arguments

`src`	The name of the query bucket.
	biocOSN Bioc's Open Storage Network (NSF) OSN bucket (spatialdata v0.3.0, zarr v2)
	sandbox scverse's spatialdata-sandbox bucket at EMBL.

Value

a vector of example (zipped) SpatialData stores available at `source`

Examples

```
Sys.setenv(AWS_REGION = "us-east-1")
library(paws.storage)

SD.data_available()
SD.data_available("sandbox")
SD.data_available("biocOSN")
```

SD.data_list *SD.data_list*

Description

Returns metadata of available data from Bioc OSN and scverse spatialdata- sandbox S3 buckets

Usage

```
SD.data_list(metadata = FALSE)
```

Arguments

metadata	if TRUE, all columns will be returned, e.g. File size, License etc.

Value

a vector of dataset names or a data.frame

Examples

```
SD.data_list()
SD.data_list(metadata = TRUE)
```

SD.data_load *retrieve scverse-curated 'SpatialData' .zarr archive*

Description

This function consolidates the retrieval and caching and transformation of scverse-curated Zarr archives and 10x-curated Xenium archives.

Usage

```
SD.data_load(id, target = tempfile(), src)
```

Arguments

id	character string; dataset identifier
target	character(1), defaults to tempfile(); use a different value if you wish to retain the unzipped .zarr store persistently.
src	The name of the source, i.e. query bucket.
	biocOSN Bioc's Open Storage Network (NSF) OSN bucket (spatialdata v0.3.0, zarr v2)
	sandbox scverse's spatialdata-sandbox bucket at EMBL.

Details

- MouseIntestineVisHD: Visium HD 3.0.0 (10x Genomics) dataset of mouse intestine; source (biocOSN): https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-librarie
- MouseBrainVisHD: Visium HD 4.0.1 (10x Genomics) dataset of mouse brain; source (sandbox): https://www.10xgenomics.com/datasets/visium-hd-three-prime-mouse-brain-fresh-frozen
- MouseBrainVis: Visium (10x Genomics) dataset of mouse brain; source (sandbox): https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-11114
- LungAdenocarcinomaMCMICRO: MCMICRO dataset of human small cell lung adenocarcinoma; source (biocOSN)
- MouseBrainMERFISH: MERFISH dataset of mouse brain tissue; source (biocOSN)
- MouseLiverMERFISH: MERFISH dataset of mouse liver tissue (SPArrOW output); source (sandbox): https://www.biorxiv.org/content/10.1101/2024.07.04.601829v1
- MulticancerSteinbock: imaging mass cytometry dataset of four cancers; source (biocOSN): https://www.nature.com/articles/s41596-023-00881-0
- ColorectalCarcinomaMIBITOF: MIBI-TOF dataset of colorectal carcinoma; source (biocOSN)
- JanesickBreastVisiumEnh: Visium (10x Genomics) dataset of breast cancer; source (biocOSN): https://www.nature.com/articles/s41467-023-43458-x
- JanesickBreastXeniumRep1: first of two Xenium (10x Genomics) sections associated with the Visium section from Janesick *et al.*; source (biocOSN)
- JanesickBreastXeniumRep2: second of two Xenium (10x Genomics) sections associated with the Visium section from Janesick *et al.*; source (biocOSN)
- HumanLungMulti_10x: Xenium (10x Genomics) data on lung cancer; source (biocOSN): https://www.10xgenomics.com/datasets/preview-data-ffpe-human-lung-cancer-with-xenium-multim
- SpaceMHelaniH3T3: SpaceM on Hepa and NIH3T3 cells; source (sandbox); more info: https://github.com/giovp/spatialdata-sandbox/blob/main/spacem_helanih3t3/README.md

Value

an instance of SpatialData, or throws an error if the dataset identifier (id)does not uniquely match the name of any resource, see SD.data_list for the list of datasets.

Examples

```
Sys.setenv(AWS_REGION = "us-east-1")

# load using `SD.data_load`
ld <- SD.data_load("ColorectalCarcinomaMIBITOF")
ld

# load data from biocOSN source (Zarr v2)
ld <- SD.data_load("ColorectalCarcinomaMIBITOF", src = "biocOSN")
```

Index

SD.data_available, 2
SD.data_list, 3, *4*
SD.data_load, 3