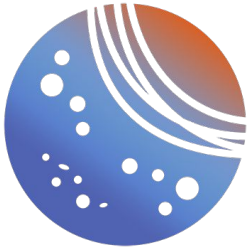


HATS (Hierarchical Adaptive Tiling Scheme) in IVOA Universe

Neven Caplar and Melissa DeLucchi (LINCC Frameworks)

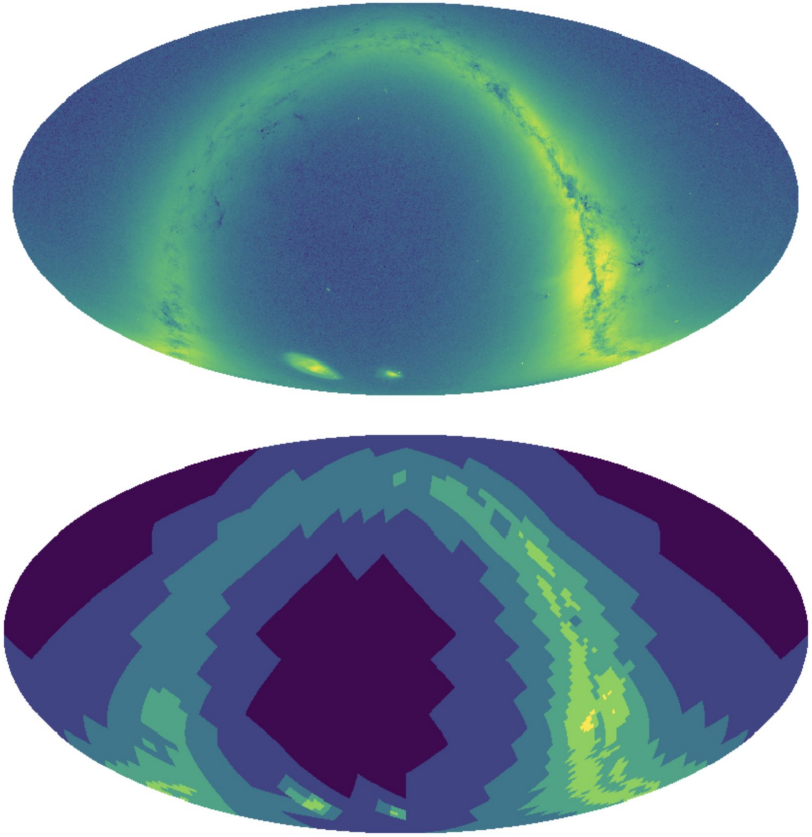
In collaboration with:

LSST Discovery Alliance

Schmidt Sciences

Short introduction of the format

- HATS (Hierarchical Adaptive Tiling Scheme):
  - Data is stored in a hierarchical data storage scheme, where the sky is split into HEALPix tiles until each tile has roughly a similar number of objects (rows).
  - These tiles are stored as Parquet files within a directory tree that encodes their location on the sky.



Drivers of development

- Previous versions: LSD, AXS

Drivers of development

- Previous versions: LSD, AXS
- LINCC Frameworks start in late 2022
  - With the mission to enable science from Vera C. Rubin Observatory
  - Parallel with development of LSDB - python & DASK based code which natively uses HATS
  - Enabling time-domain science through Nested package

Drivers of development

- Previous versions: LSD, AXS
- LINCC Frameworks start in late 2022
  - With the mission to enable science from Vera C. Rubin Observatory
  - Parallel with development of LSDB - python & DASK based code which natively uses HATS
  - Enabling time-domain science through Nested package
- Monthly working group meeting in Americas-friendly time zone & Europe-friendly time zone
- Busy week August 2024
- HATS-on-the-fly, LSDB server, Firefly-over-LSDB

Structure - data leafs

```
catalog/
|-- [REQUIRED] hats.properties
|-- [RECOMMENDED] partition_info.csv
|-- [OPTIONAL] skymap.fits
|-- [OPTIONAL] data_thumbnail.parquet
+-- dataset/
    |-- [RECOMMENDED] _metadata
    |-- [RECOMMENDED] _common_metadata
    |-- Norder=0/
    |-- Norder=1/
    |-- Norder=2/
    |-- Norder=3/
    |-- Norder=4/
    |-- Norder=5/
    |-- Norder=6/
    +-- Norder=. . ./
```

Listing 1: Example catalog directory contents

- Parquet files with recommended _healpix_29 column
- _healpix_29 can repeat

- Data stored at appropriate level given the density at that spatial location (and only at that level)
- Final leaf directories allow for addition of small amounts of data without having to recreate the catalog
- Useful for alerts and daily updates

```
dataset/
|-- . . .
|-- Norder=6/
|   |-- Dir=0/
|   |   |-- Npix=0/
|   |   |   |-- part0.parquet
|   |   |   +-- . . .
|   |   +-- Npix=9999/
|   |       |-- part0.parquet
|   |       +-- . . .
|   +-- Dir=10000/
|       +-- Npix=10000/
|           |-- part0.parquet
|           +-- . . .
|-- Norder=7/
+-- . . .
```

Listing 3: Example catalog dataset directory contents with leaf directories

Structure - _healpix_29

- Parquet files with recommended _healpix_29 column
- _healpix_29 can repeat

```
euclid
```

✓ 0.2s

lsdb Catalog euclid_q1_merFinalCatalog:

	OBJECT_ID	RIGHT_ASCENSION	DECLINATION	RIGHT_ASCENSION_PSF_FITTING
npartitions=85				
Order: 4, Pixel: 637	int64[pyarrow]	double[pyarrow]	double[pyarrow]	double[pyarrow]
Order: 5, Pixel: 2557	...	...	...	...
...	...	...	...	...
Order: 4, Pixel: 2256	...	...	...	...
Order: 4, Pixel: 2258	...	...	...	...

472 out of 472 columns in the catalog have been loaded **lazily**, meaning no data has been read, only the catalog schema

```
euclid.head(5)
```

✓ 1.5s

_healpix_29	OBJECT_ID	RIGHT_ASCENSION	DECLINATION	RIGHT_ASCENSION_PSF_FITTING
718324103215191225	2699943881634300724	269.994388	63.430072	<NA>
718324103729508935	2699935797634310616	269.99358	63.431062	<NA>
718324105113206067	2699890168634331416	269.989017	63.433142	<NA>
718324105820999925	2699917121634334052	269.991712	63.433405	<NA>
718324105824218050	2699920771634334559	269.992077	63.433456	<NA>

5 rows × 472 columns

Structure - catalog collections

```
gaia_dr3/
|-- collection.properties
|-- catalog/
|   |-- hats.properties
|   +-- . . .
|-- [OPTIONAL] margin_5arcs/
|   |-- hats.properties
|   +-- . . .
|-- [OPTIONAL] margin_10arcs/
|   |-- hats.properties
|   +-- . . .
+-- [OPTIONAL] index_designation/
    |-- hats.properties
    +-- . . .
```

Listing 5: Example collection directory contents

- Allows to have auxiliary catalogs connected with primary catalog;
- Avoids having to explicitly load like you have to do now

Auxiliary files and metadata

- [**REQUIRED**] **hats.properties**, at catalog level / Required file defining HATS catalog metadata
- [**RECOMMENDED**] **partition_info.csv**, at catalog level / Lists HEALPix tiles used for partitioning
- [**RECOMMENDED**] **partition_join_info.csv**, at catalog level / Maps primary to join catalog partitions
- [**OPTIONAL**] **skymap.fits**, at catalog level / HEALPix skymap showing sky coverage per tile
- [**OPTIONAL**] **data_thumbnail.parquet**, at catalog level / Small, representative sample of dataset
- [**RECOMMENDED**] **_metadata**, at catalog/dataset level / Aggregated Parquet footers with statistics
- [**RECOMMENDED**] **_common_metadata**, at catalog/dataset level / Dataset-wide Parquet schema
- [**REQUIRED**] **collection.properties**, at catalog collection level / Required metadata file for catalog collections

Summary of Feedback

Bold green = Completed

IVOA associations

Connection between HiPS and HATS

How will VOParquet and HATS work together

Be more precise about MOC/HEALPix skymap

Precisions and clarifications

Be more precise in footer and header definitions - difference between Parquet and metadata footer/header

Index tables description should be more exact

Rename properties to hats.properties [WIP]

Other metadata files apart from _metadata and _common_metadata [WIP]

skymap.fits flavors [WIP]

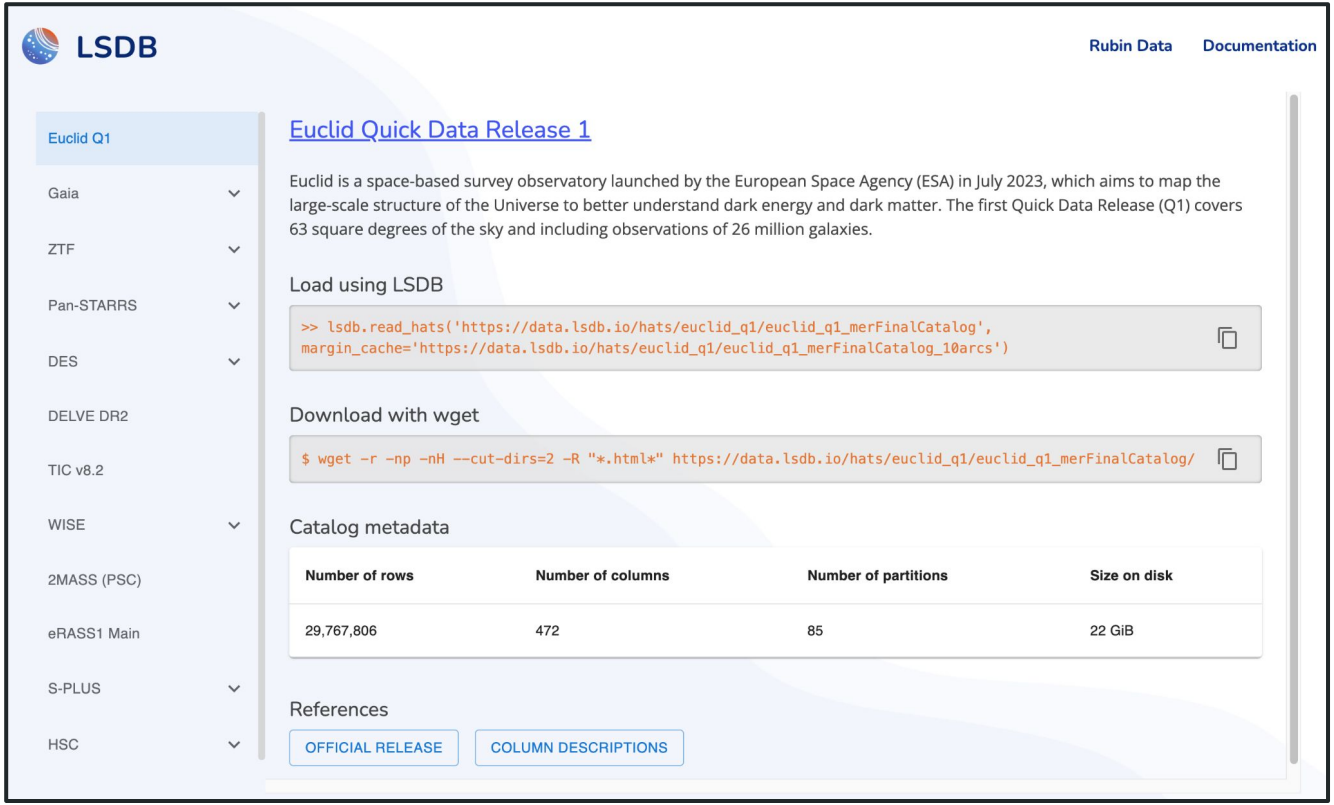
Allow storage of parquet leaf files not as files on disk, but as a list of files that map to each partition.

Enhancements

Current status

data.lsdb.io

Serving public catalogs

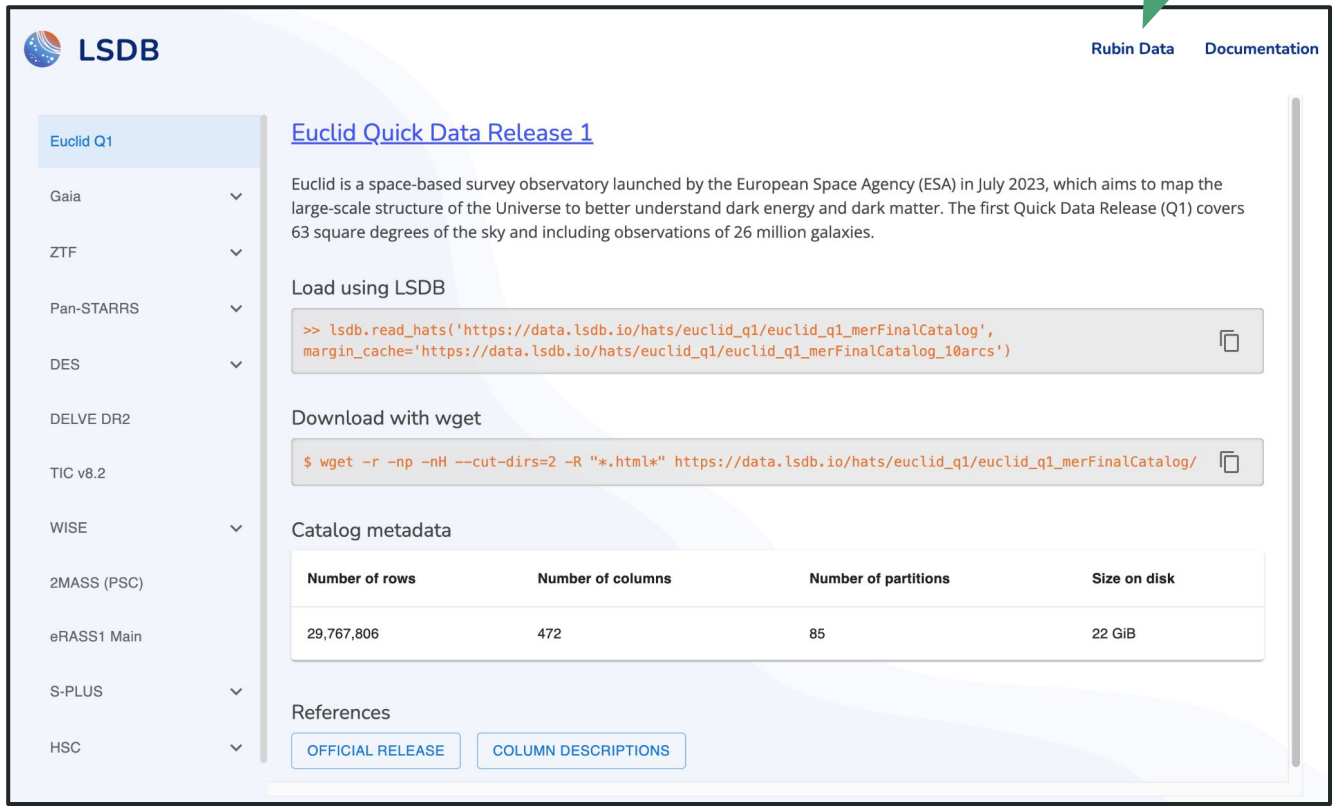


Current status

Rubin Data

data.lsdb.io

Serving public catalogs



Future

- Serving catalogs with CDS (next talk by FX)
- Further Collaboration
  - IPAC (ZTF, Euclid, Fornax)
  - Space Telescope (Roman, PanSTARRS, Fornax)
- Official status with Vera C. Rubin before LSST Data Release 1

- Using HATS for other types of astronomical data
  - Collaboration with HEASARC to provide Fermi data
  - Collaboration with NOIRLab and DESI to provide spectral data (and photometric redshifts)

- Collecting feedback and user experience with the goal of updating the note and towards IVOA standardization. Determining data discovery plan for HATS.

Future

https://docs.lsdb.io/en/latest/contact.html
Slack channel: **#lincc-frameworks-lsdb**
@discovery-alliance.slack.com
Working group**:** **hats-wg@googlegroups.com**
10am Pacific, 3rd Friday of month
6am Pacific, 4th Thursday of month
Office hours:
10am Pacific, every Friday
10am Pacific, every Thursday (LINCC F.)

- Serving catalogs with CDS (next talk by FX)
- Further Collaboration
  - IPAC (ZTF, Euclid, Fornax)
  - Space Telescope (Roman, PanSTARRS, Fornax)
- Official status with Vera C. Rubin before LSST Data Release 1

IVOA note here

- Using HATS for other types of astronomical data
  - Collaboration with HEASARC to provide Fermi data
  - Collaboration with NOIRLab and DESI to provide spectral data (and photometric redshifts)
- Collecting feedback and user experience with the goal of updating the note and towards IVOA standardization. Determining data discovery plan for HATS.