A Science Discovery Portal for Euclid
Euclid Scientific Archive System

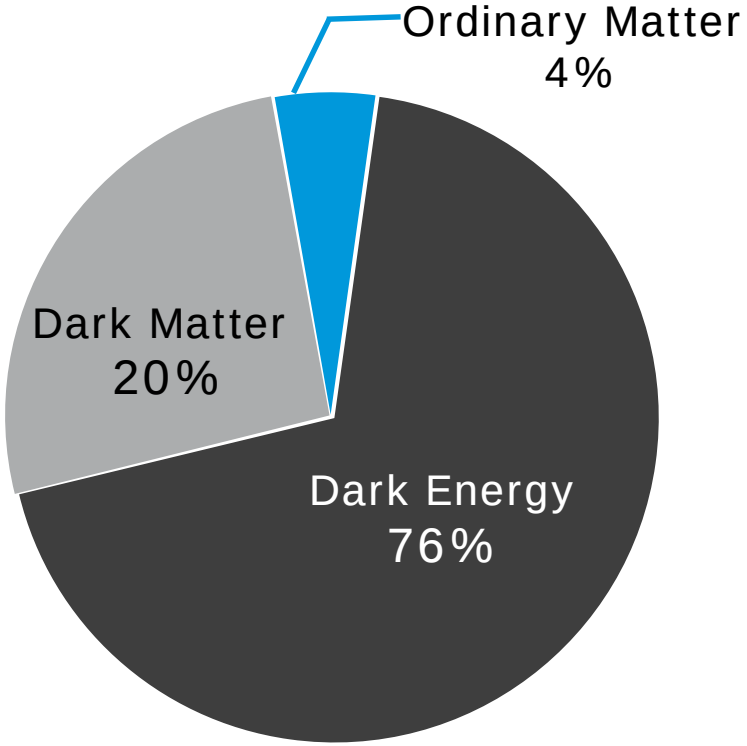
Sara Nieto on behalf of

ESAC Science Data Centre (ESDC)

Euclid Science Operations Centre (SOC)

Euclid Goals

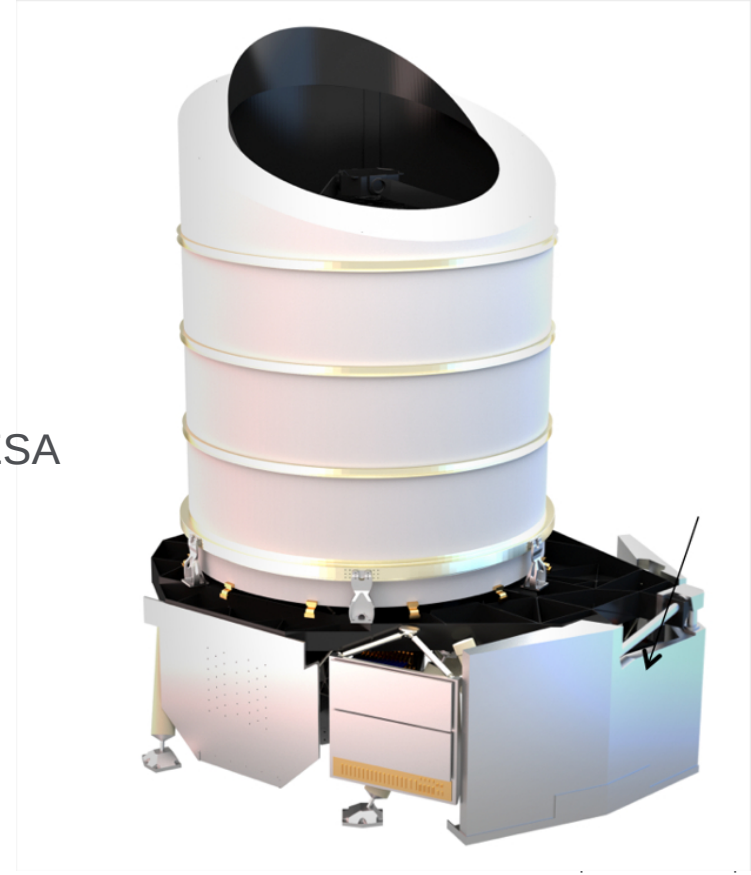
- Discover the origin of the Universe's accelerating expansion.
- Discover the nature of 95% of the Universe: dark energy and dark matter.
- Measure shapes of galaxies distorted by gravitational deflection due to dark matter.
- Measure non-random distribution of galaxies resulting from the action of gravity



Euclid Mission Overview

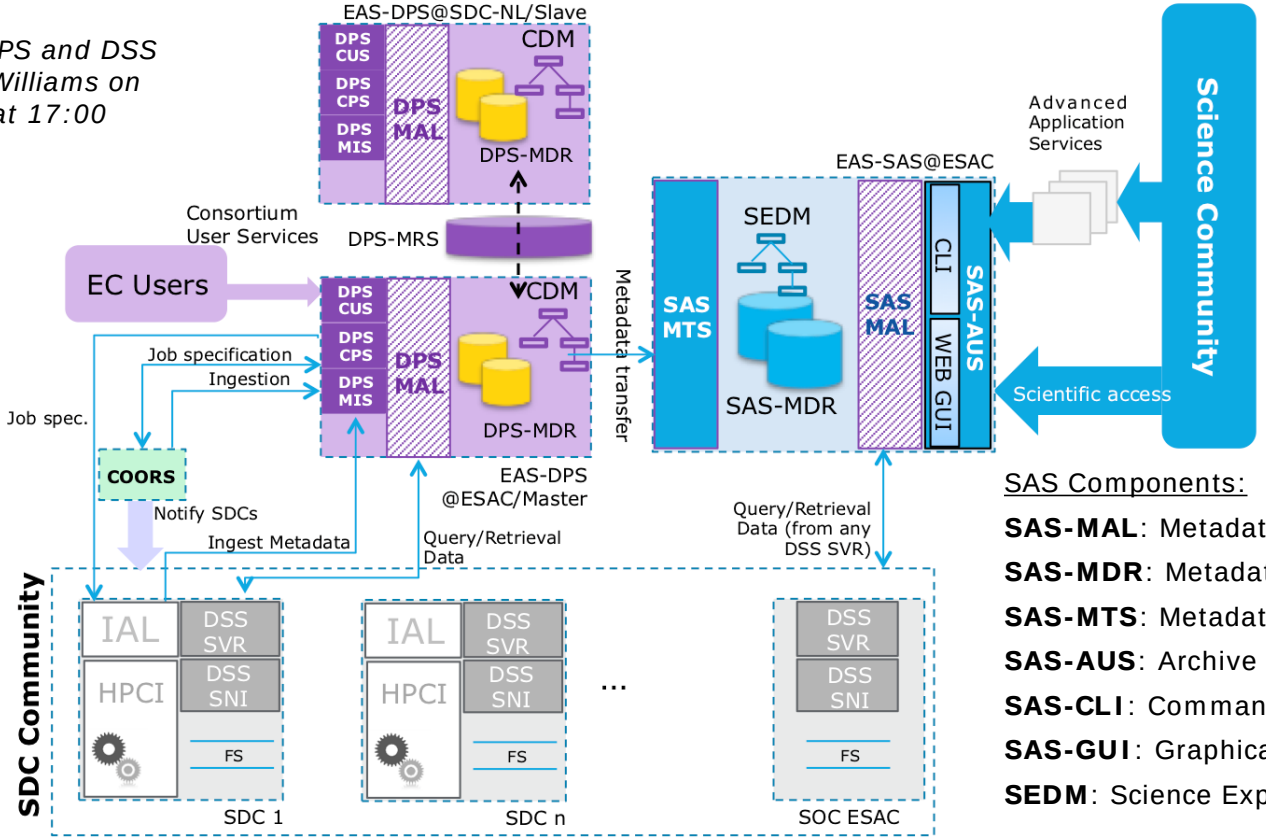
- **Launch in Q2 2022** and 6 years duration
- 1.2m telescope, L2 orbit
- Visible-light camera and a near-infrared camera/spectrometer

- **ESA** is responsible for the mission
- The **Euclid Consortium** (EC) will supply ESA with the instruments and most of the SGS
- Euclid Consortium & Other teams
  - 15 countries, 130 institutes, 1300 consortium members



Euclid Archive System Overall Architecture

Talk on DPS and DSS
By Rees Williams on
Tuesday at 17:00



SAS Components:

SAS-MAL: Metadata Access Service

SAS-MDR: Metadata Repository

SAS-MTS: Metadata Transfer Service

SAS-AUS: Archive User Services

SAS-CLI: Command Line Interface

SAS-GUI: Graphical User Interface

SEDM: Science Exploitation Data Model

Euclid Science

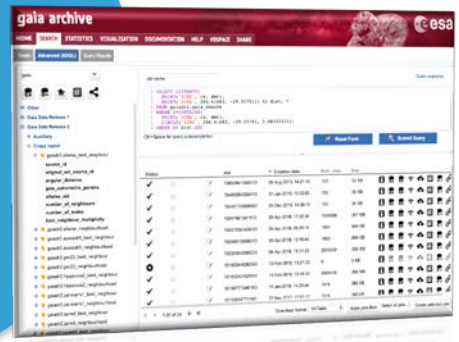
sky.esa.int

Images

LE3

Catalogues

Spectra

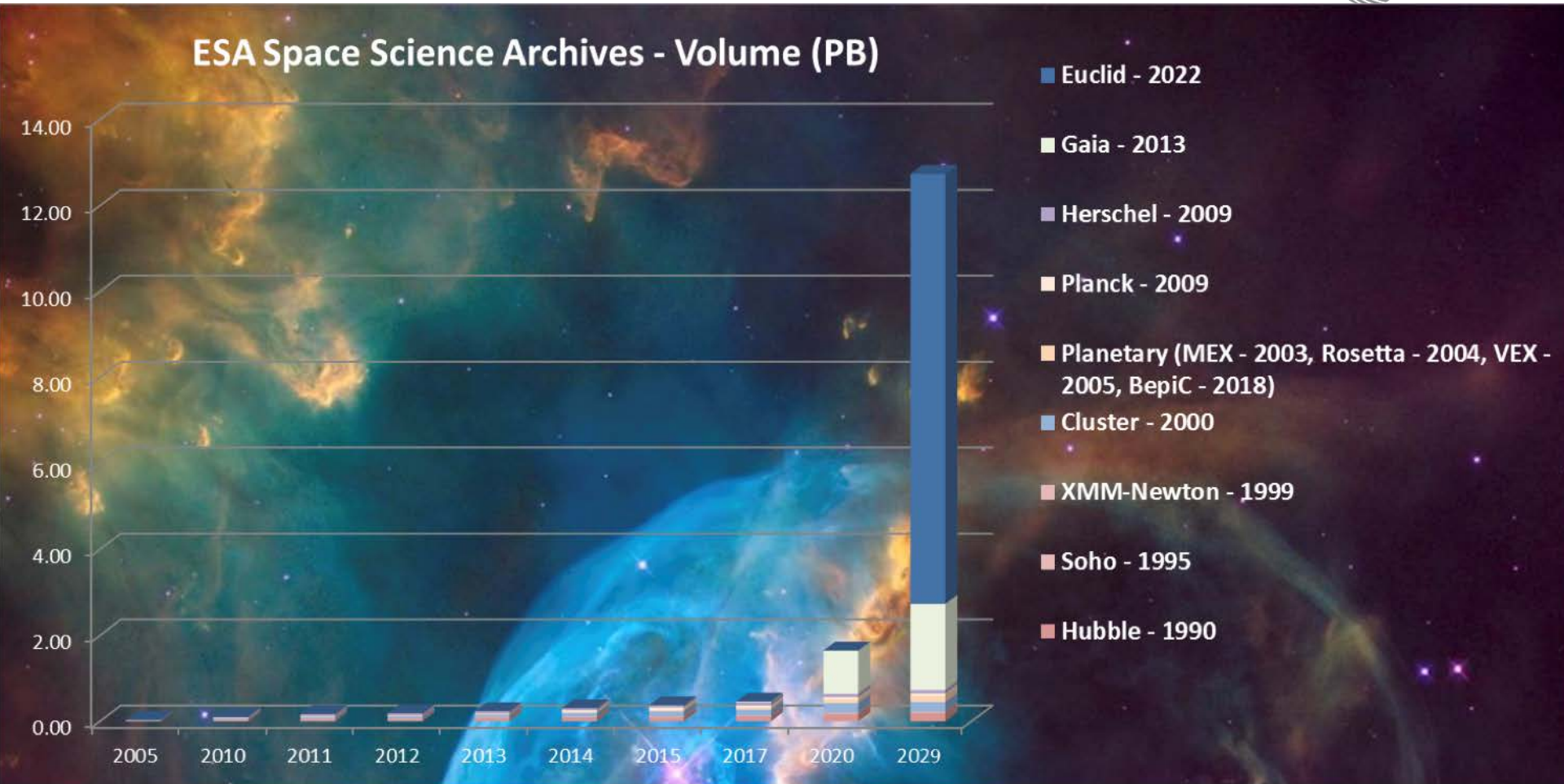


gea.esac.esa.int

nxsa.esac.esa.int

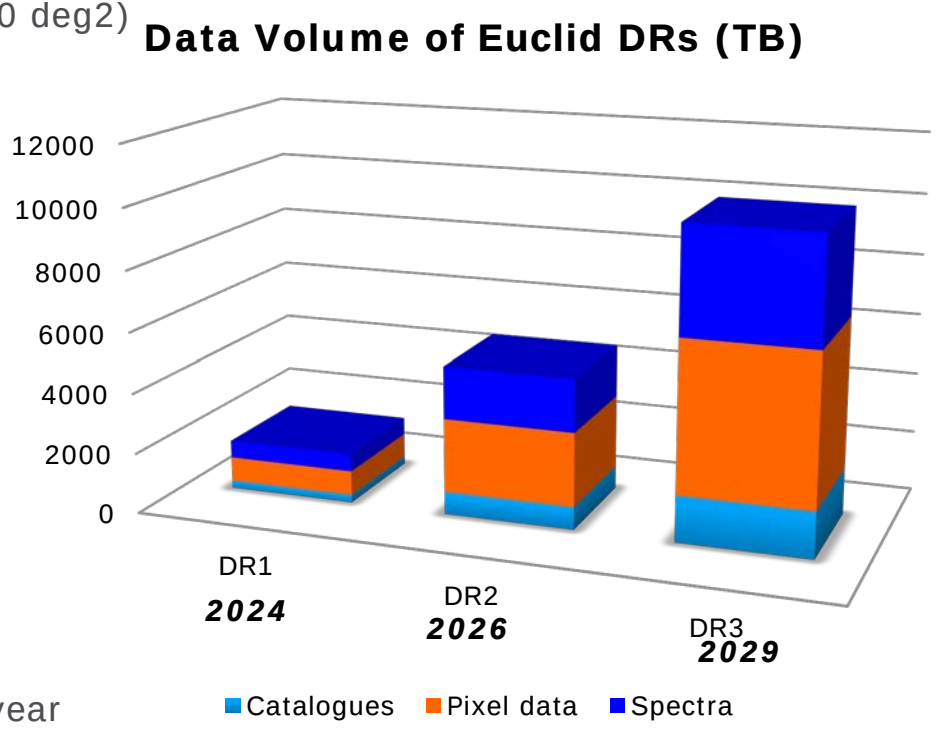
Talk on Gaia datasets by Juan Gonzalez on Wednesday at 13:45

ESA Science Archives Volume Evolution (2000-2030)



Euclid Science Volume and Data Releases

- **10PB of data**, ~45000 observations in 6 years mission
- Wide survey (15000 deg2) and Deep survey (40 deg2)
  - **Catalogue**: ~1.50PB
    - MER: 50TB
    - SPE columns: 244TB
    - PHZ columns: 188TB
    - SHE columns: 1PB
  - **Imaging**: ~5.5PB
    - VIS & NIR: 3.5PB
    - MER: 2PB
  - **Spectra**: ~3.2PB
  - Other archive products, HiPS maps
  - External data (3-5PB) KiDS, etc.
- Mission data processing could generate 26 PB/year



The Euclid Science Archive System

Scientific Community

WEB Portal

Science Exploitation Platform

VO Compliant APIs

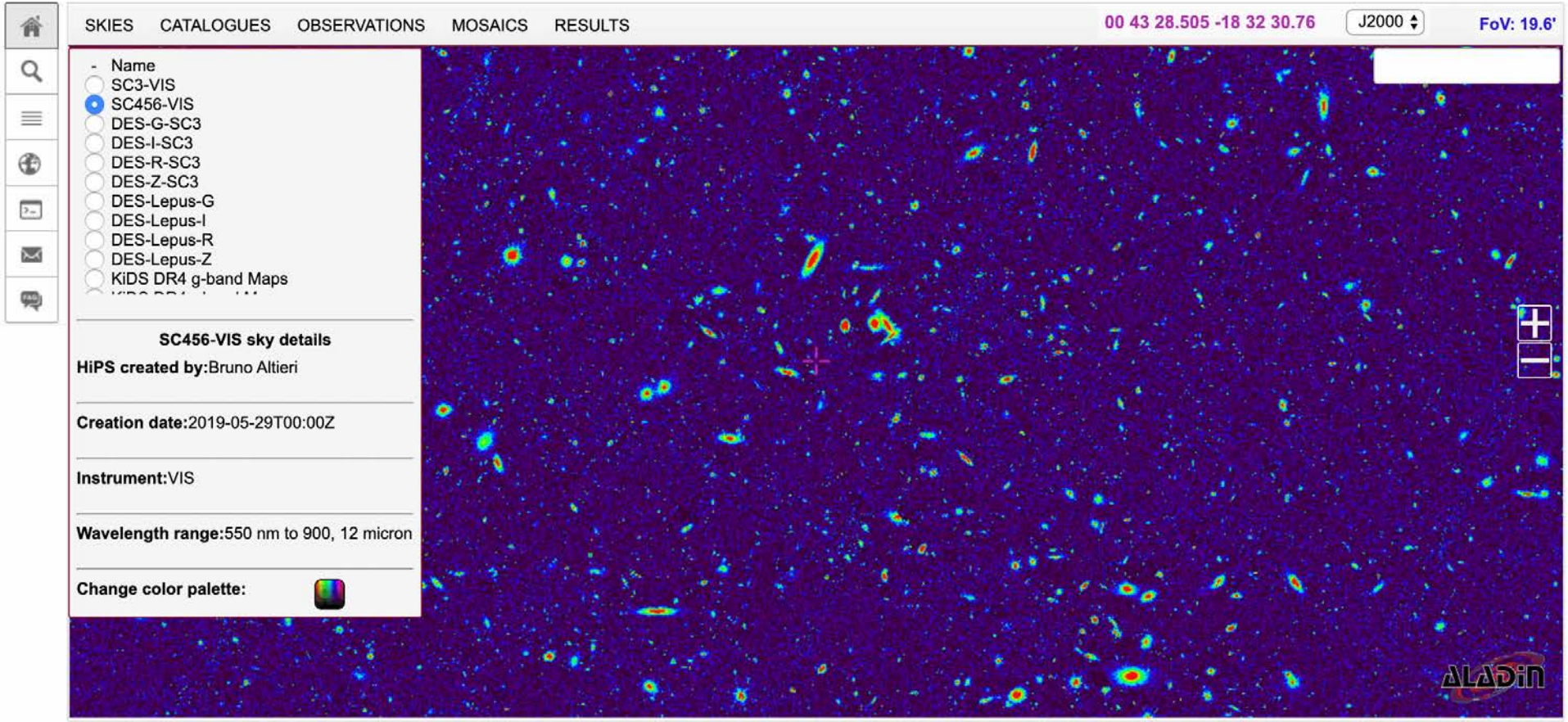
TAP+ | DataLink | SIA | SSA | SODA | VOSpace

Greenplum open source
MPP DB

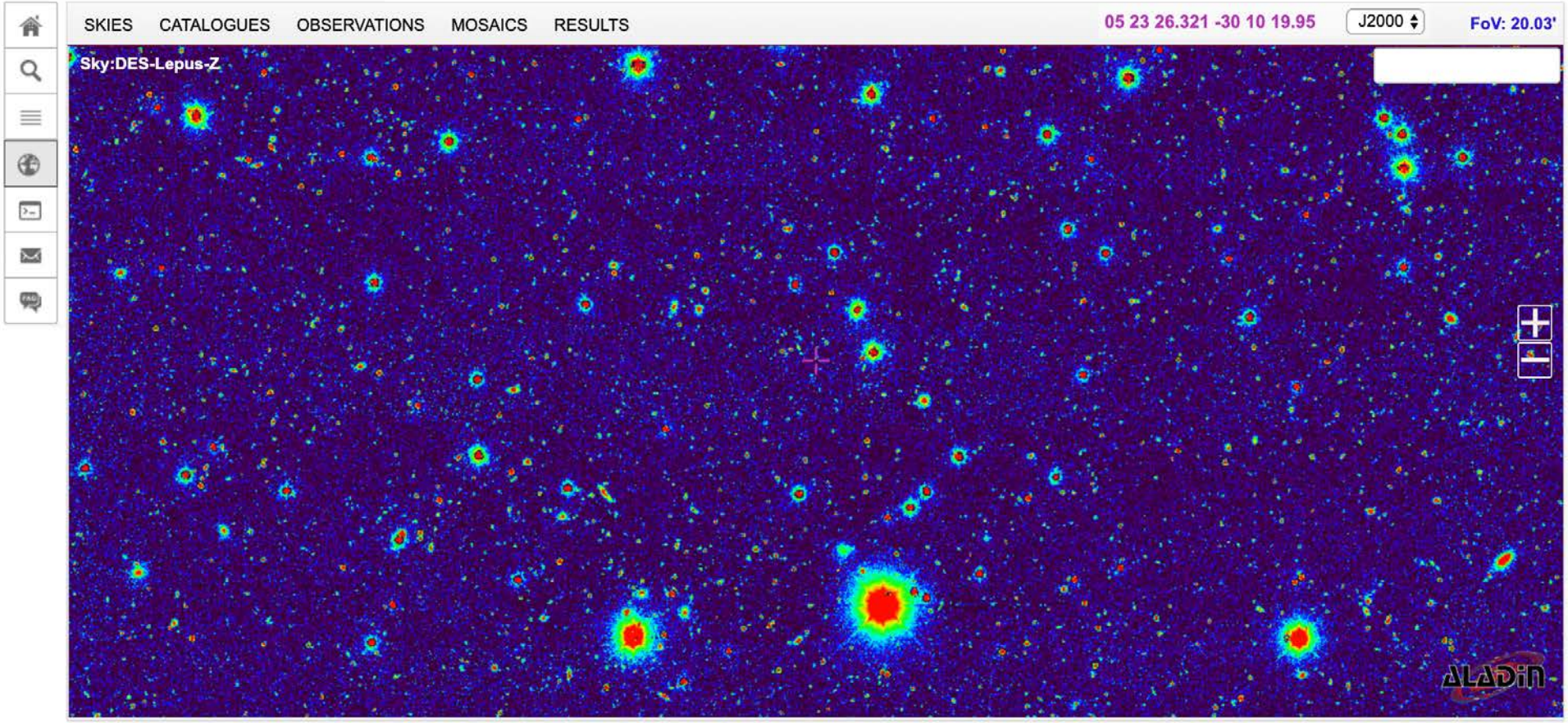
Petascale Data Storage
ceph

Ingestion Layer

SAS v0.9: Euclid HiPSized simulations



SAS v0.9: External data, DES HiPS maps



SAS v0.9: Browse simulated observations

RESULTS #2 ✖ RESULTS #3 ✖

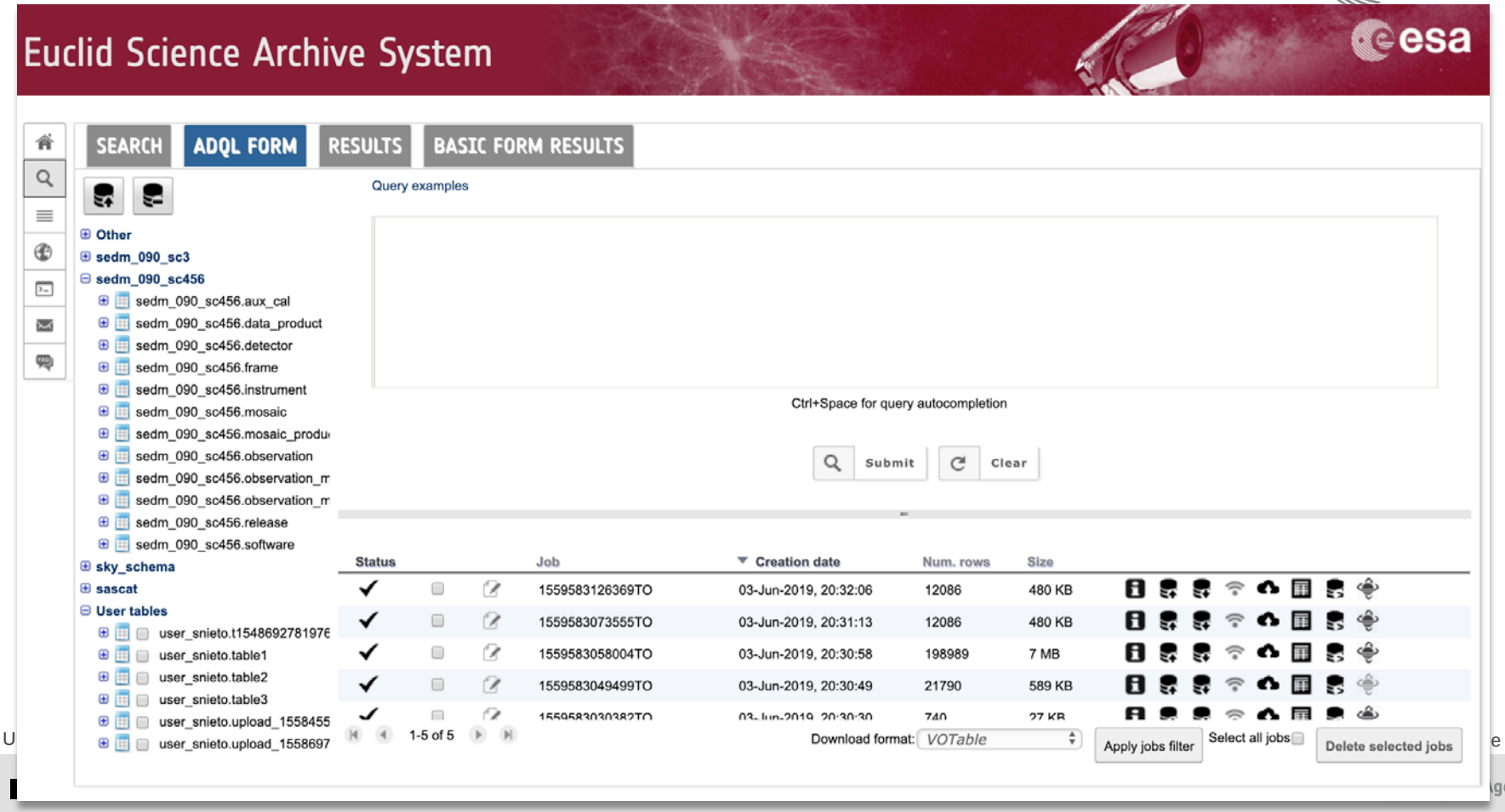
observations (27) mosaics (210)

Observation ID		RA	Dec	Instrument	Filter	Observing Mode	Start Time
52926		00h 35m 16.58s	-19d 28' 29.48"	NISP	NIR_H	SCIENCE WIDE	2025-06-22T18:56:13.0
52926		00h 35m 16.58s	-19d 28' 29.48"	NISP	NIR_J	SCIENCE WIDE	2025-06-22T18:56:13.0
52926		00h 35m 16.58s	-19d 28' 29.48"	NISP	NIR_Y	SCIENCE WIDE	2025-06-22T18:56:13.0
53401		00h 29m 51.54s	-20d 04' 09.15"	NISP	NIR_J	SCIENCE WIDE	2025-06-21T13:34:03.0
53401		00h 29m 51.54s	-20d 04' 09.15"	NISP	NIR_Y	SCIENCE WIDE	2025-06-21T13:34:03.0
53401		00h 29m 51.54s	-20d 04' 09.15"	NISP	NIR_H	SCIENCE WIDE	2025-06-21T13:34:03.0
53402		00h 32m 34.18s	-19d 46' 18.41"	NISP	NIR_J	SCIENCE WIDE	2025-06-21T17:14:03.0
53402		00h 32m 34.18s	-19d 46' 18.41"	NISP	NIR_H	SCIENCE WIDE	2025-06-21T17:14:03.0
53402		00h 32m 34.18s	-19d 46' 18.41"	NISP	NIR_Y	SCIENCE WIDE	2025-06-21T17:14:03.0
53403		00h 36m 38.51s	-20d 09' 47.42"	NISP	NIR_Y	SCIENCE WIDE	2025-06-22T17:42:53.0
53403		00h 36m 38.51s	-20d 09' 47.42"	NISP	NIR_J	SCIENCE WIDE	2025-06-22T17:42:53.0
53403		00h 36m 38.51s	-20d 09' 47.42"	NISP	NIR_H	SCIENCE WIDE	2025-06-22T17:42:53.0

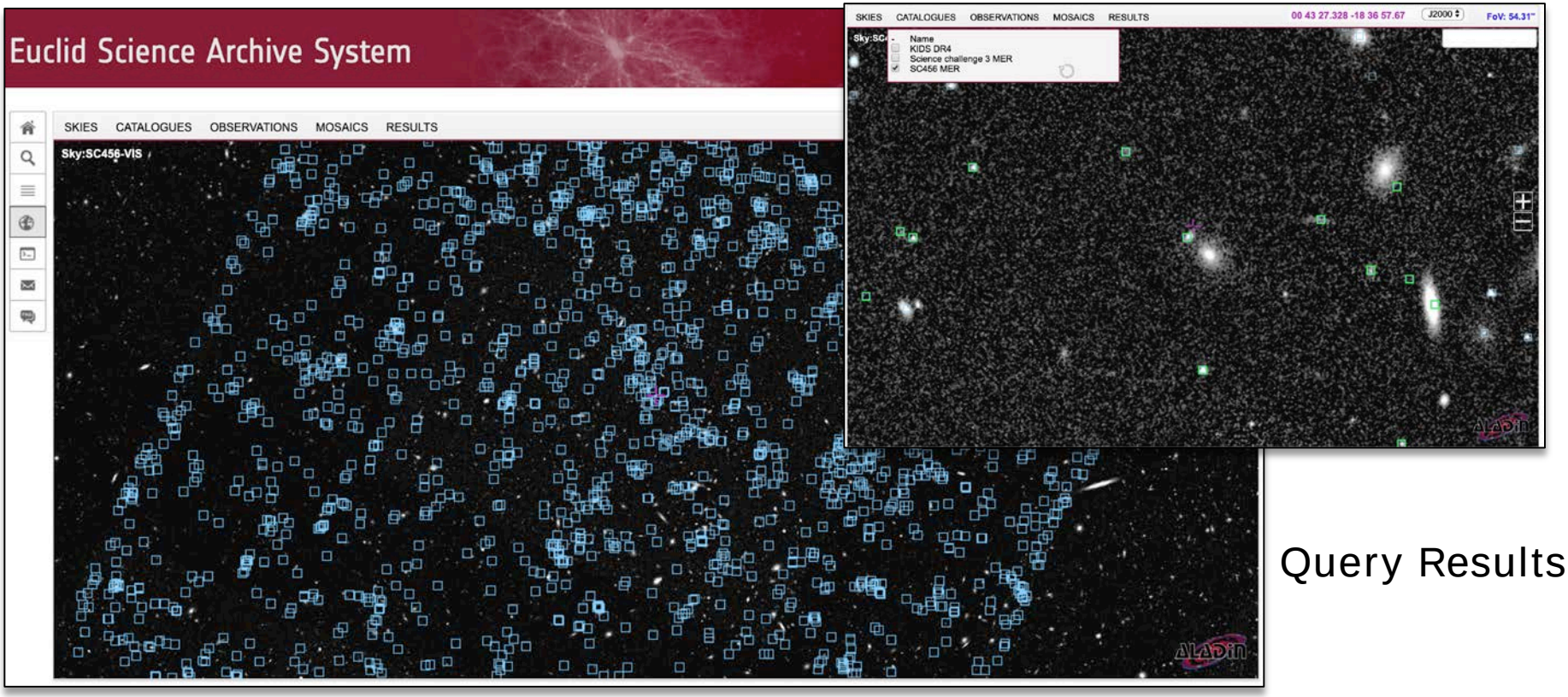
1 of 1 Page size: 50

Displaying 1-27 of 27

SAS v0.9: ADQL Query Interface

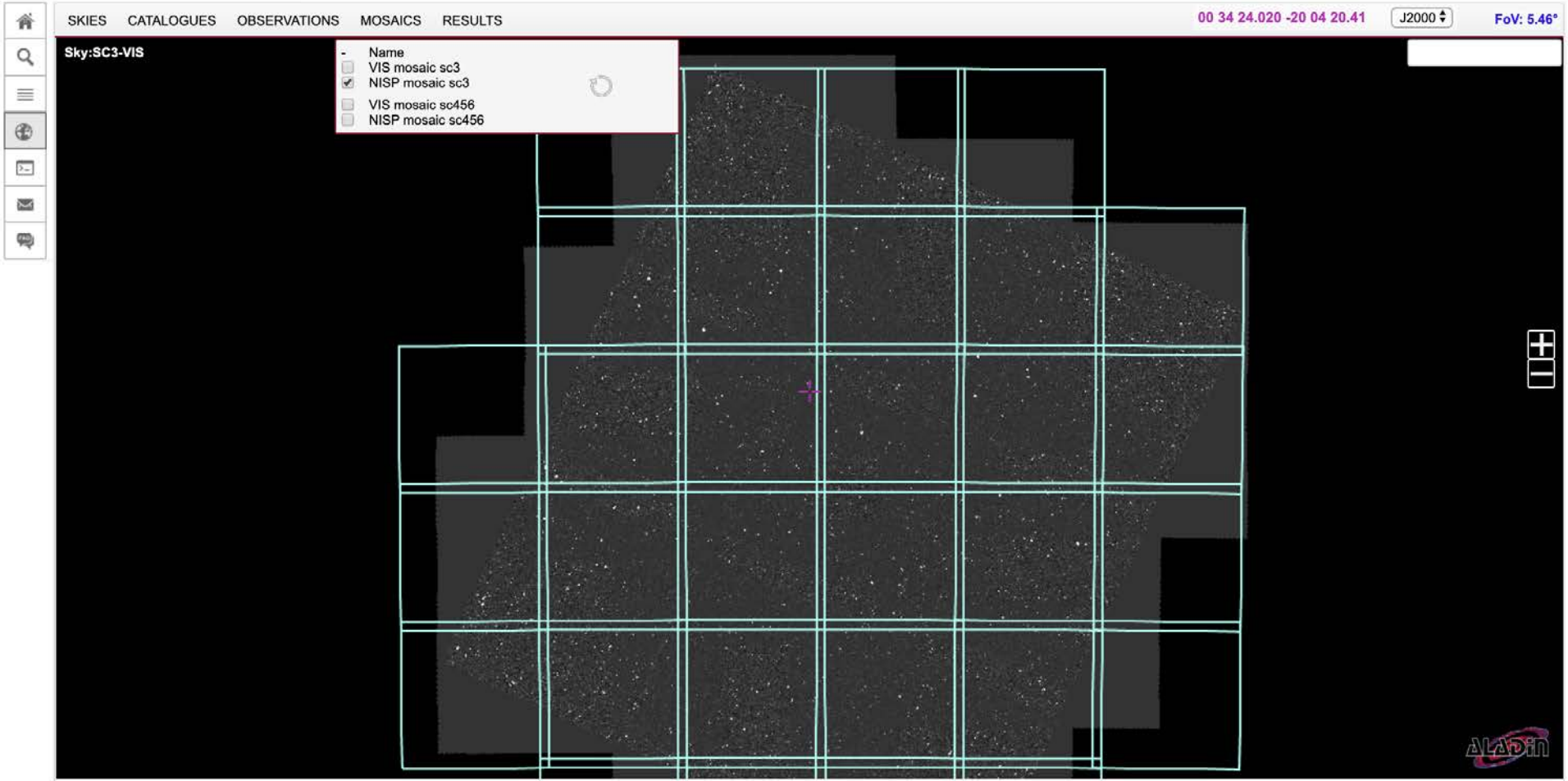


SAS Web Portal v0.9: Overlay of query results

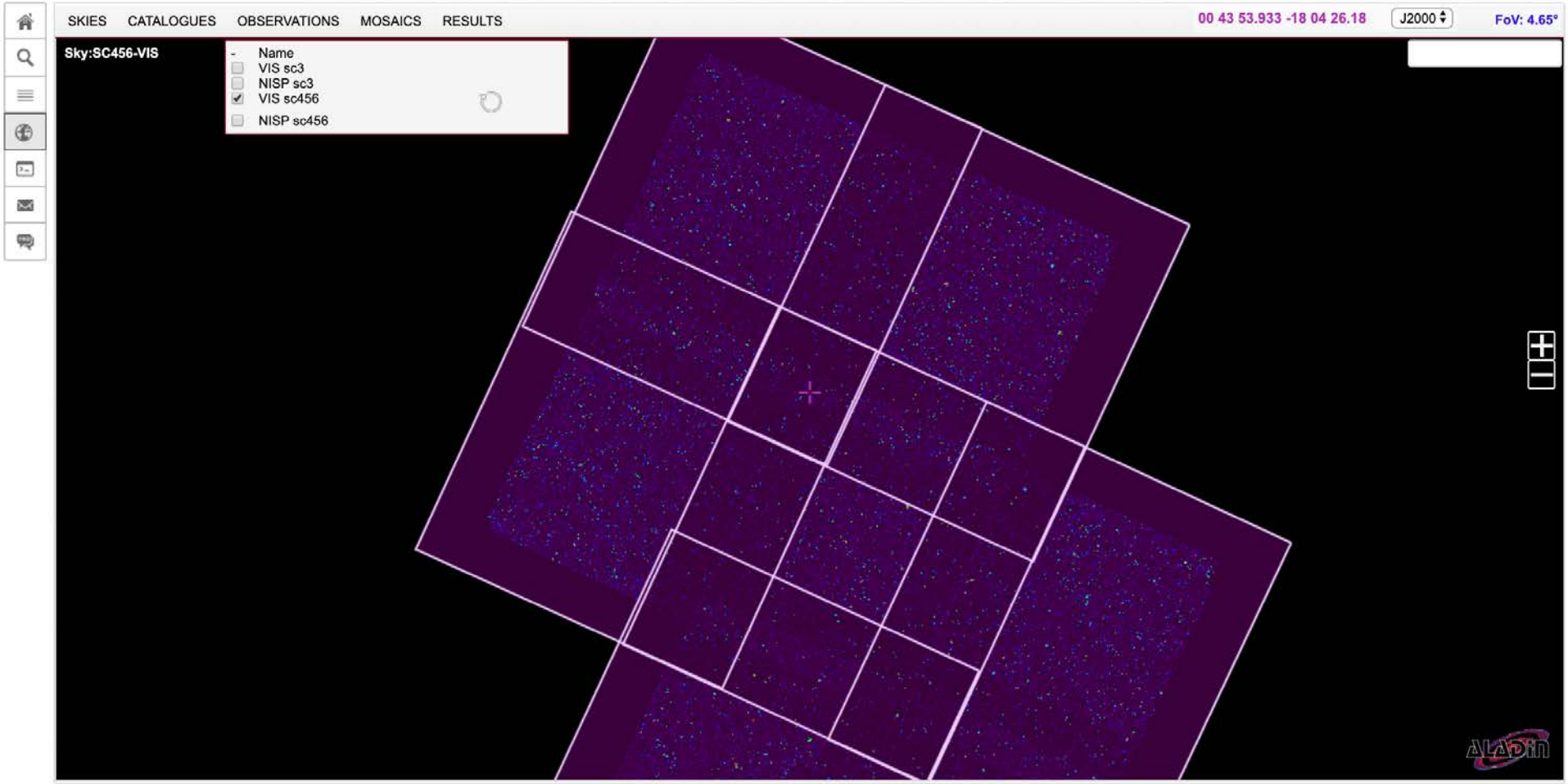


Query Results

SAS Web Portal v0.9: Mosaics



SAS Web Portal v0.9: Footprints of observations



SAS Data, Service and Client Layers

DATA

Catalogues
- EUC Flagship 2.7B
- KiDS DR#4 100M
- DES Lepus 3.5M

Images and HiPS maps
- SC#3
- DES, DES Lepus
- KiDS DR#4

SC#456 new data

Footprints
- Observations
- Mosaics

Services

- EuclidSky
- Overlay of sources
- Overlay of user results

- Catalogue searches
- Metadata searches
- Download service

- **Cutout of images**
- **Plotting service**

- Spectra
- DPS Interface
- SEPP interface
- X-match

Client

WEB Portal
- EuclidSky explorer
  - HiPS maps
  - Catalogues
  - Footprints

- Products browser

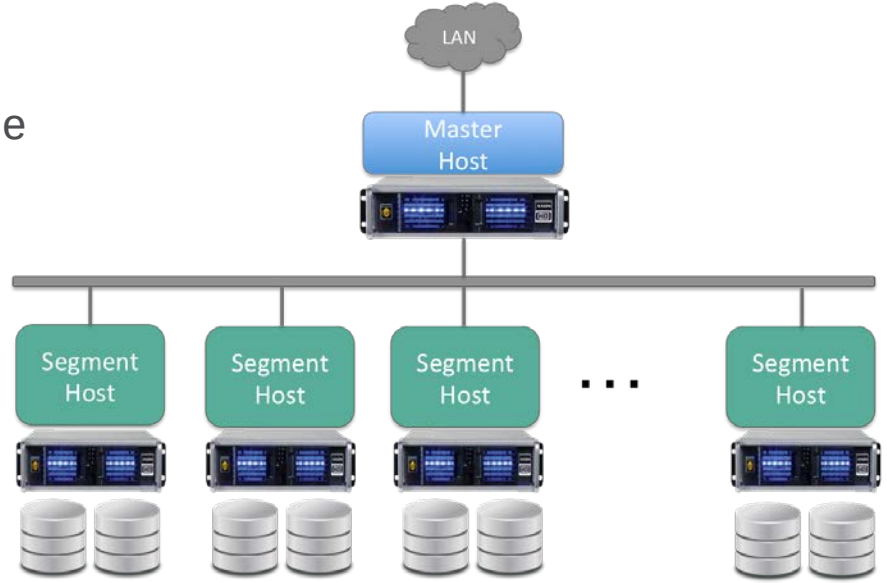
- **Plotting browser**

- Spectra visualizer
- AstroQuery modules
  - TAP+
  - Cutout
- JupyterLab

SAS
V0.10

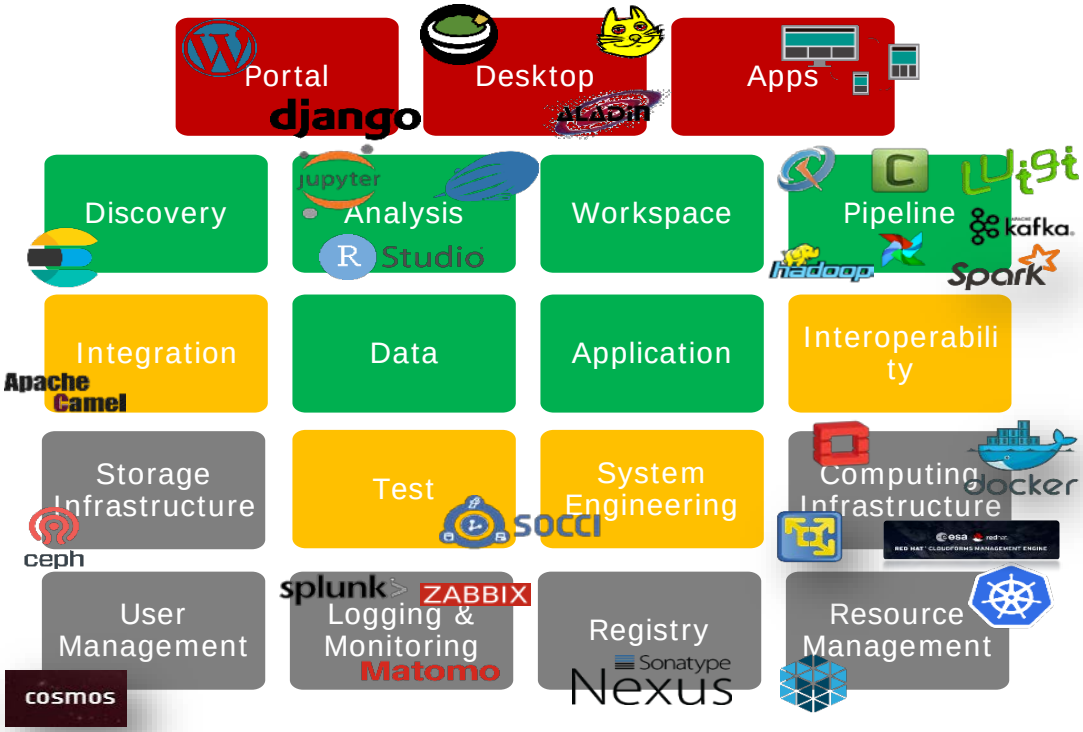
Petascale Infrastructure

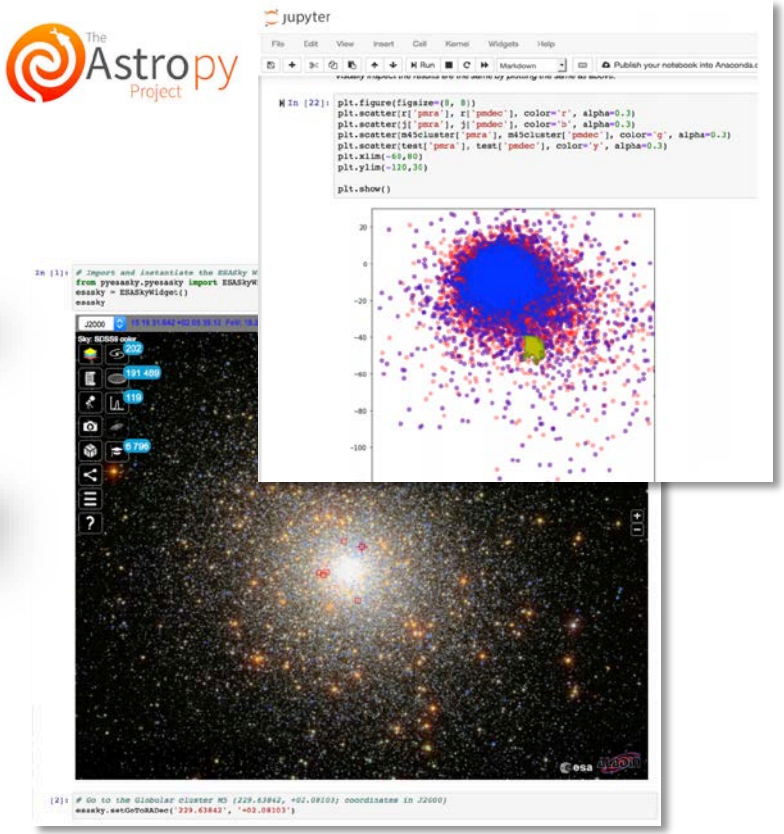
- **GreenPlum** open source now used as DB system in SAS replacing PostgreSQL
- 10 VM Nodes of 128GB Master and 32GB Workers with 2TB size/node NetApp
  - *Catalogue of ~3 Billion sources,~ 100 columns*
- Current NetApp NFS storage
- **Ceph** PoC for scaling out ESDC storage
  - Massive scalable (to Exa-Bytes)
  - Highly reliable
  - Easy to manage
  - Open source
- **Apache Spark** PoC



Move code to the data: ESA SEPP

Talk By Vicente Navarro on Tuesday at 16:30





Questions

Thanks for your attention

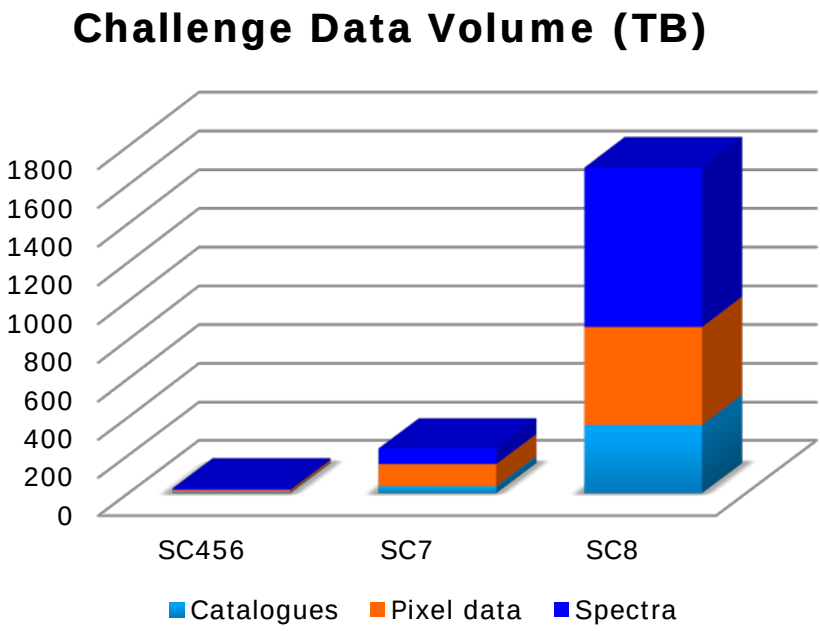
Additional information

Science Archive participation in Euclid Challenges

- The Scientific Archive is officially participating Euclid Challenges since #3

- Challenges roadmap up to launch in 2022
  - 3 deg2 (9 fields)
  - **30 deg2 (~109 fields) on going**
  - 300 deg2 wide (~1000 fields)
  - 3000 deg2 wide (~10000 fields)

- Internal Data Releases after each challenge
- Switch of Data Processing System @ESAC as Master in 2020



Apache Spark Proof of Concept

- Spark: distributed parallel processing framework
- Use cases: Analysis of large catalogues and pixel data (source extraction) 6-
  - 6 VM workers of 32GB RAM
  - 8 cores in standalone mode
  - NetApp (NFS) storage
- Spark ecosystem >> Astronomy
  - Example: AXS (latest ADASS)

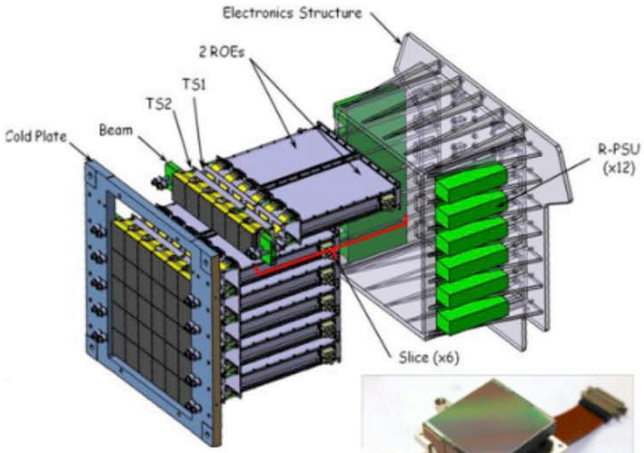
```
In [1]: fn = "/shared_spark_jupy/mer/"
        ## Create a DataFrame from the HDU data of a FITS file
        df = spark.read.format("com.astrolabsoftware.sparkfits")\
        .option("hdu", 1).load(fn).select("RightAscension","Declination","flagh","flagy","flagj","flagvi
                ,"fluxgextdecamaper","starflag","fluxyaper")
        df.printSchema()
        df.show(10)
```

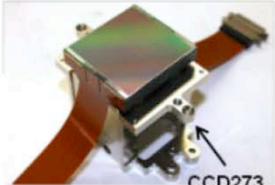
```
root
 |-- RightAscension: float (nullable = true)
 |-- Declination: float (nullable = true)
 |-- flagh: integer (nullable = true)
 |-- flagy: integer (nullable = true)
 |-- flagj: integer (nullable = true)
 |-- flagvis: integer (nullable = true)
 |-- fluxgextdecamaper: float (nullable = true)
 |-- starflag: short (nullable = true)
 |-- fluxyaper: float (nullable = true)

+---------------+------------+-----+-----+-----+-------+-----------------+--------+---------+
|RightAscension|Declination|flagh|flagy|flagj|flagvis|fluxgextdecamaper|starflag|fluxyaper|
+---------------+------------+-----+-----+-----+-------+-----------------+--------+---------+
|         8.4954| -20.418758|    0|    0|    1|      0|          36.2309|       0|187.64561|
|       8.397478| -20.419008|    0|    0|    0|      0|        17.166737|       0|108.34584|
|       8.150222| -20.418182|    1|    1|    1|      1|        84.964485|       0|601.73126|
|       8.487046| -20.417128|    1|    1|    1|      1|         994.4865|       0| 123.4418|
|       8.048966|   -20.4171|    0|    0|    0|      0|        14.058238|       0|37.825474|
|       8.387903| -20.418386|    0|    0|    0|      0|         6.383344|       0|77.042656|
|       8.100646| -20.413496|    0|    0|    0|      0|        230.17534|       0|773.78754|
|       8.029462| -20.419916|    0|    0|    0|      0|        2.0375946|       0|12.479751|
|       8.387057| -20.419903|    0|    0|    0|      0|        2.8992279|       0| 5.531452|
|       8.356397| -20.418049|    0|    0|    0|      0|        6.5645847|       0|67.025665|
+---------------+------------+-----+-----+-----+-------+-----------------+--------+---------+
only showing top 10 rows
```

VIS and NISP

Courtesy: S. Pottinger, M. Cropper and the VIS team

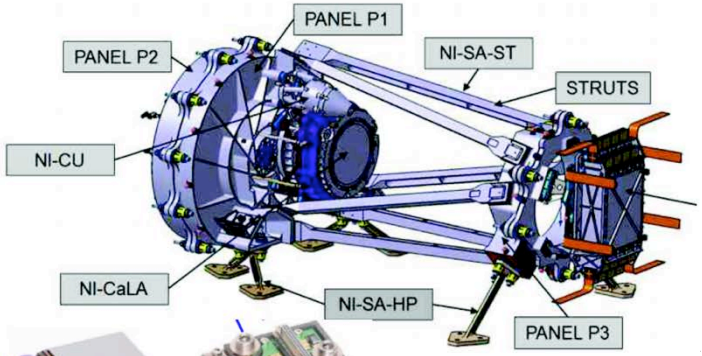
- FoV: 0.54deg^2
- Mass : 133 kg
- Telemetry: < 520 Gbt/day
- 36 4kx4K E2V CCDs, 12 micron pixels
- 0.1 arcsec pixel on sky
- Limiting mag, wide survey AB : 24.5 (10 σ)
- **1 Filter**: Y(R+I+Y): band pass 550-900nm

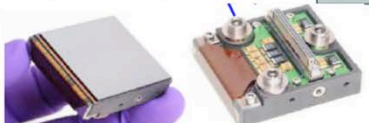




Courtesy: T. Maciaszek and the NISP team

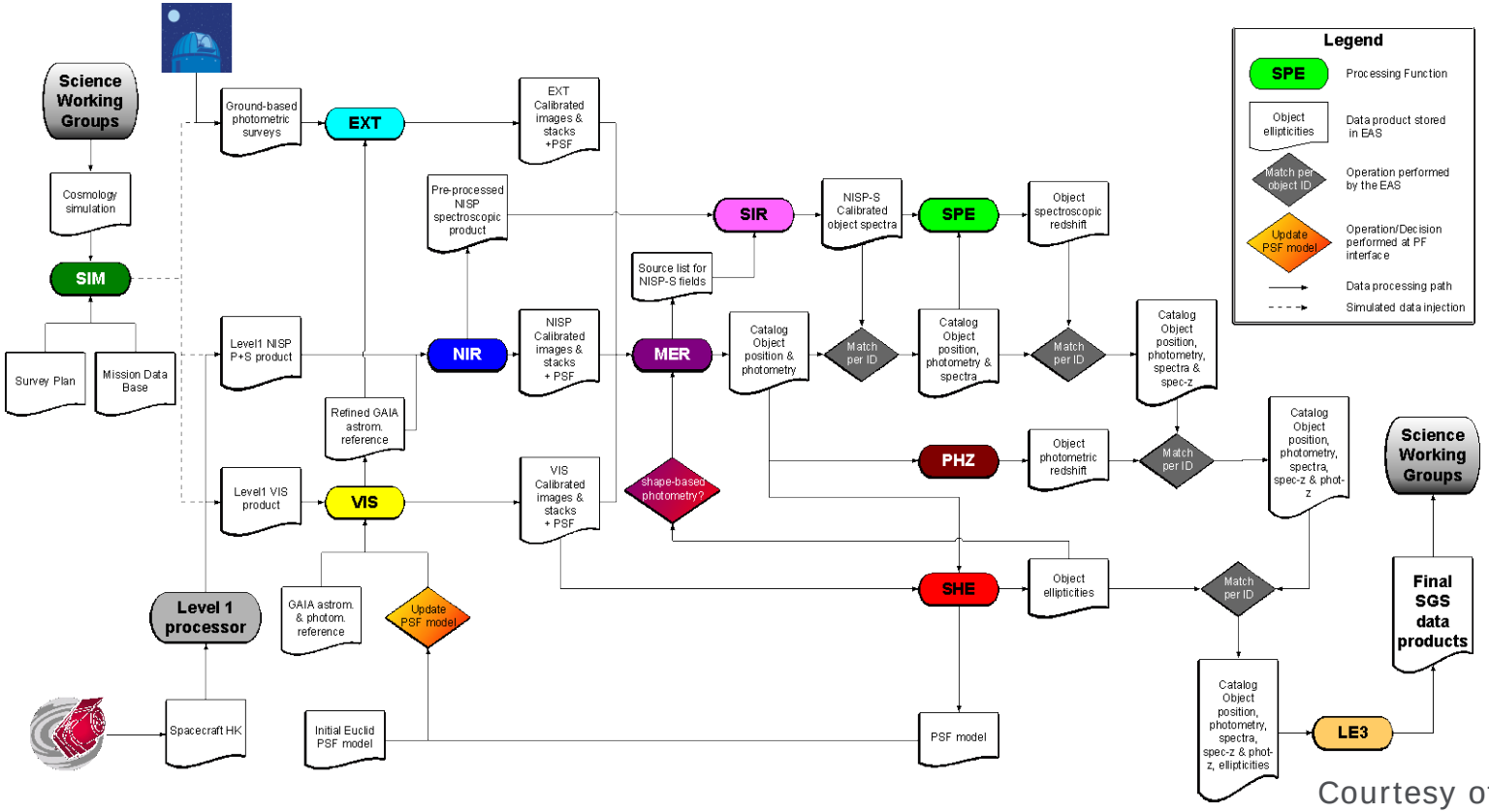
- FoV: 0.55 deg^2
- Mass : 159 kg
- Telemetry: < 290 Gbt/day
- Size: 1m x 0.5 m x 0.5 m
- 16 2kx2K H2GR detectors
- 0.3 arcsec pixel on sky
- Limiting mag, wide survey AB : 24 (5 σ)
- **3 Filters**: Y, J, H
- **4 grisms**: 1B (920 – 1250) ,3R (1250 – 1850)



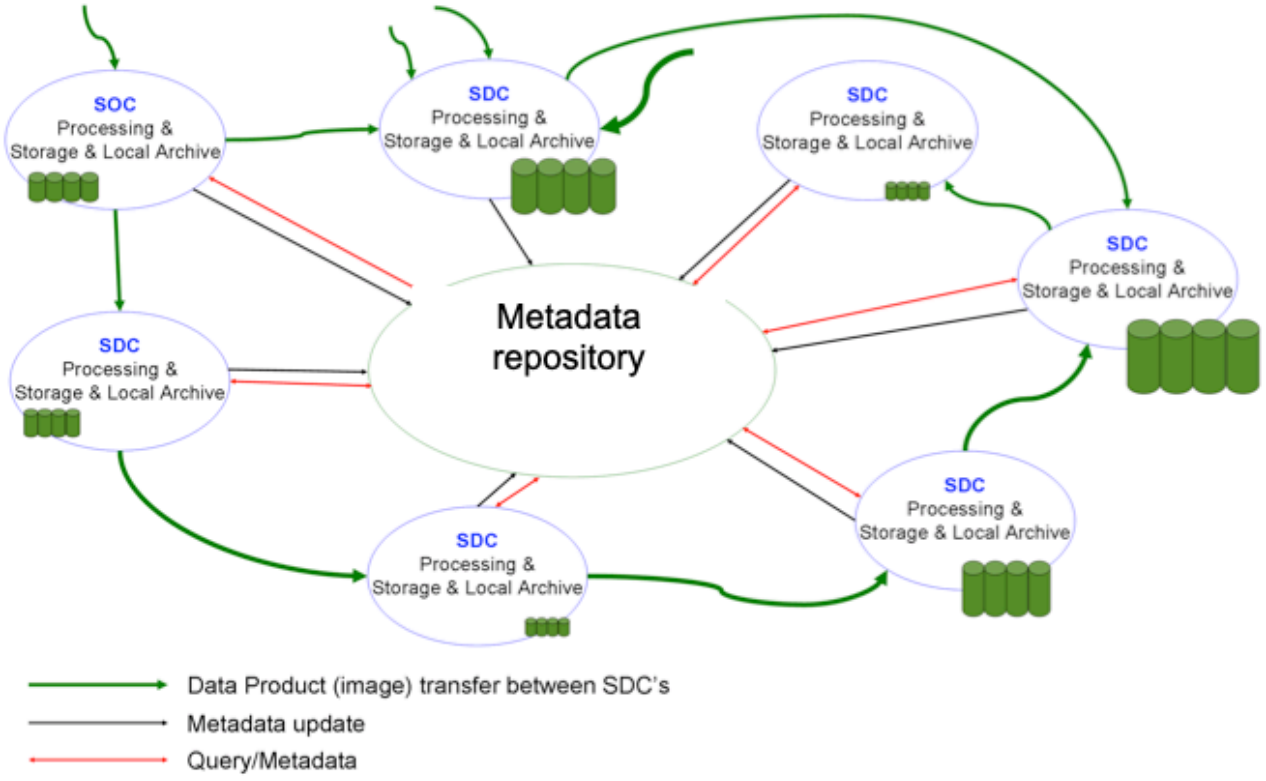


Courtesy of EC

Euclid Simulations Data Flow

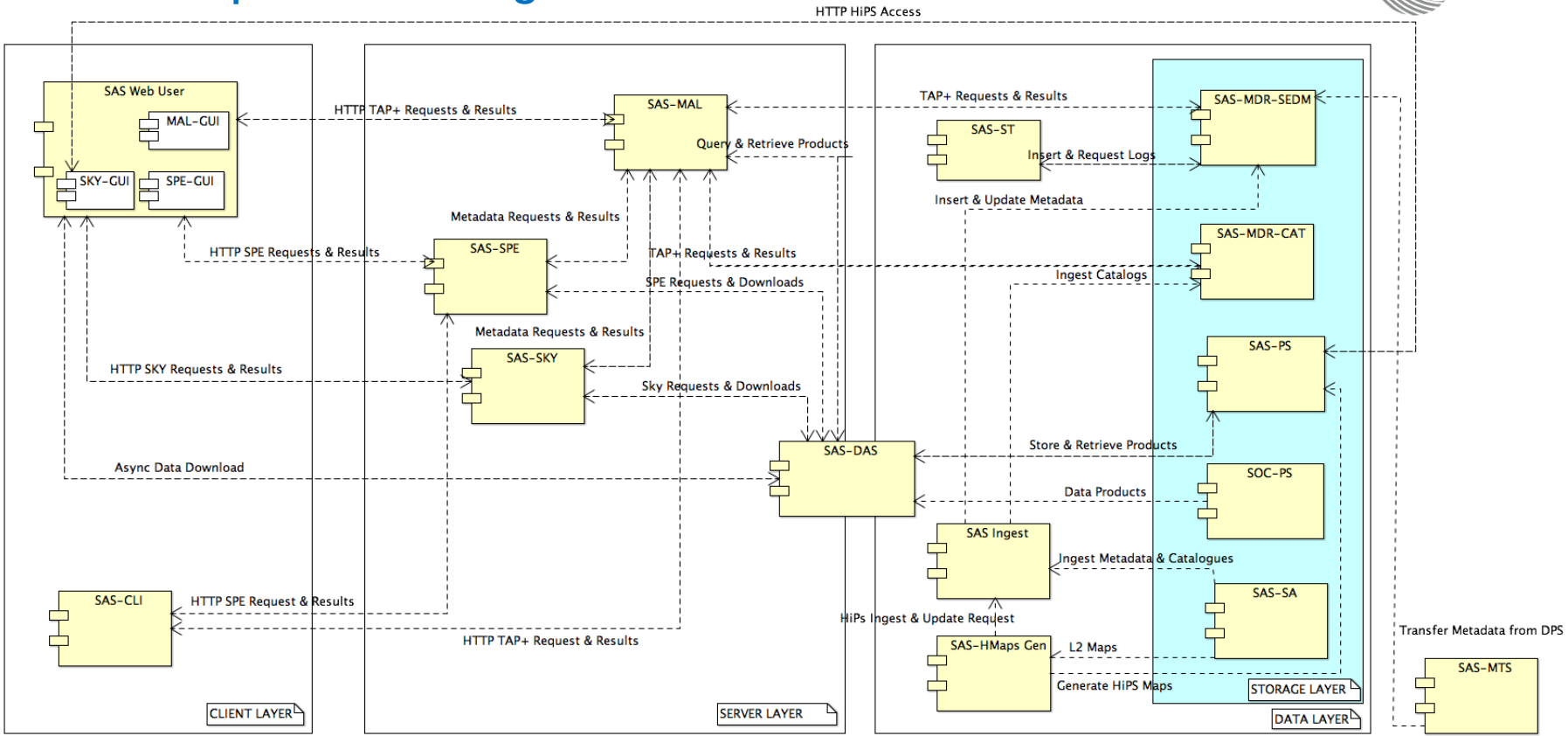


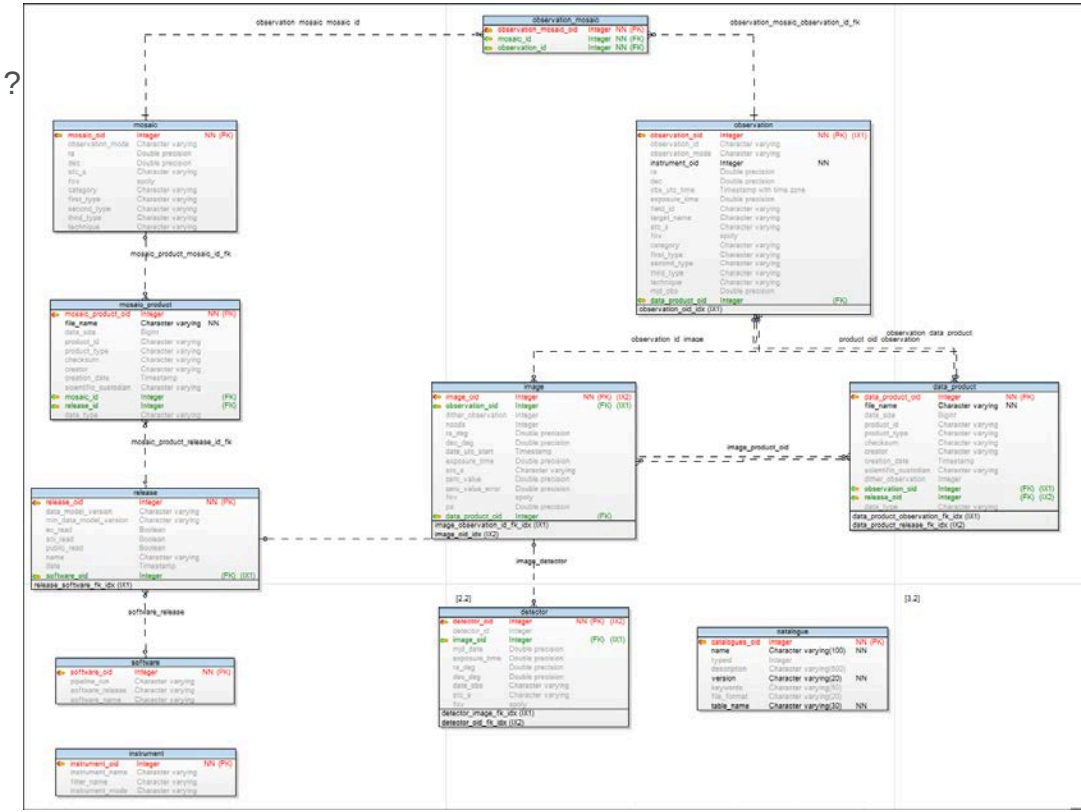
Courtesy of EC & ESA



Courtesy of EC

SAS Component Diagram



Interoperability VO Driven

- SEDM based on VODM Standards:
  - ObsCore, ProvenanceDM, SourceDM ?
- TAP+ (Table Access Protocol)
- ADQL (Astronomical Data Query Lang.)
- UWS (Universal Worker Service)
- VOSpace (Virtual Observatory space)
- HiPS (Hierarchical Progressive Survey)
- SODA (Server-side Operations Data Access)
- SIAP (Simple Image Access Prot.)
- DataLink
- Euclid SEDM evolves as of ECDM
- SEDM v0.9 is based on ECDM 1.8.2