

Data federations and its applications

Edwin A. Valentijn

Prof Astronomical Information Technology

Target- OmegaCEN – Kapteyn

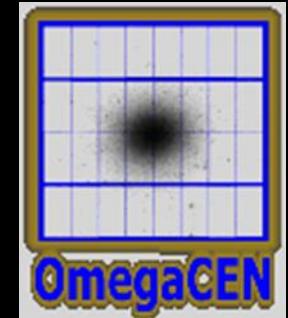


20 Oktober 2016 IAU Symp. Astro Informatics Sorrento

Astro-WISE – Data federations

Distributed Information System handling surveys

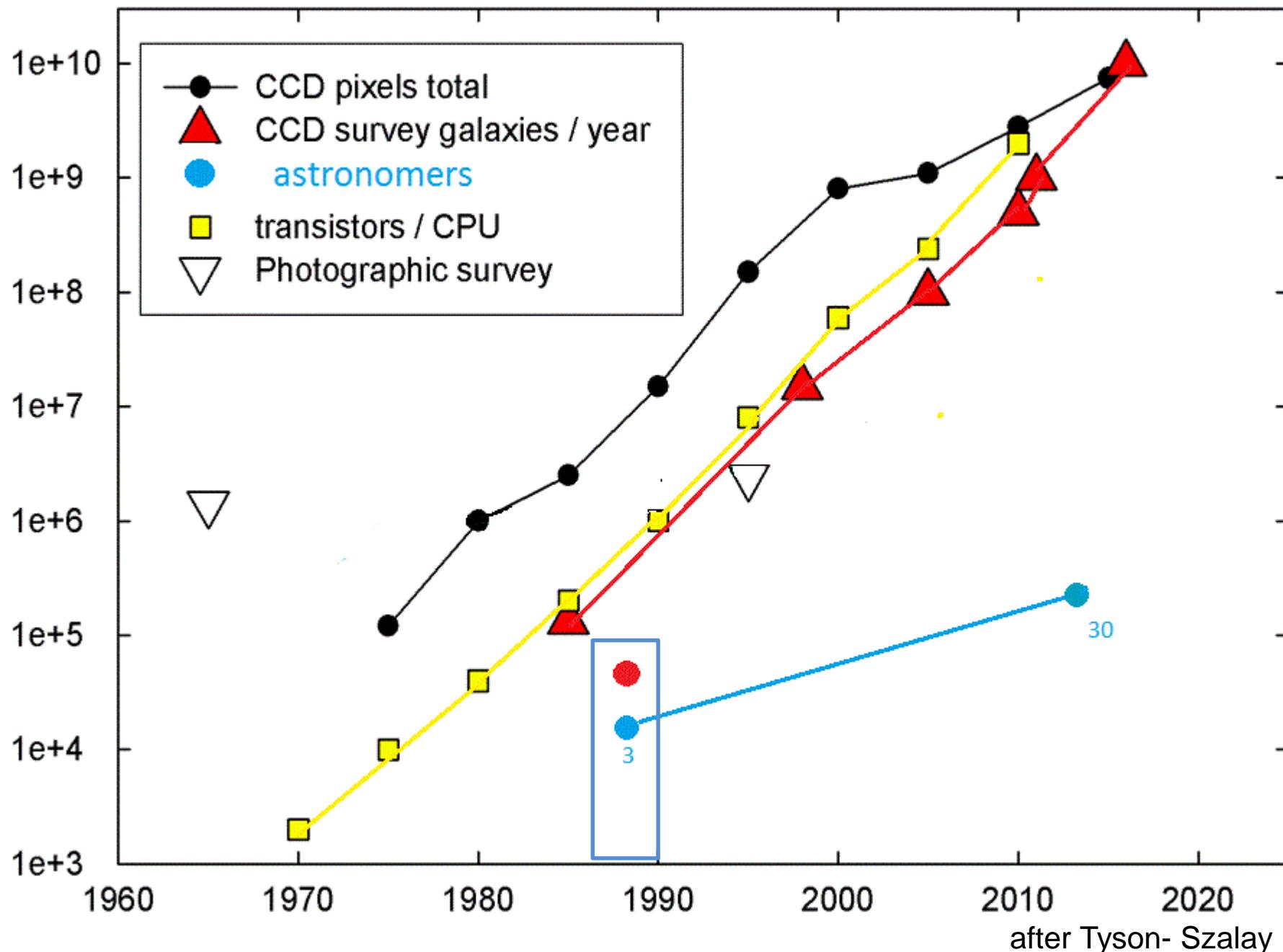
since 2003 - it works



- KiDS DR1 2 3 - ESO - VST
- MUSE - ESO - VLT
- Lofar - Astron
- -> Euclid - ESA Micado - ESO-ELT
- Glimps- AI Handwritten text – Lifelines DNA
- Target Holding

=====→ Data federations

Trends in Optical Astronomy Survey Data

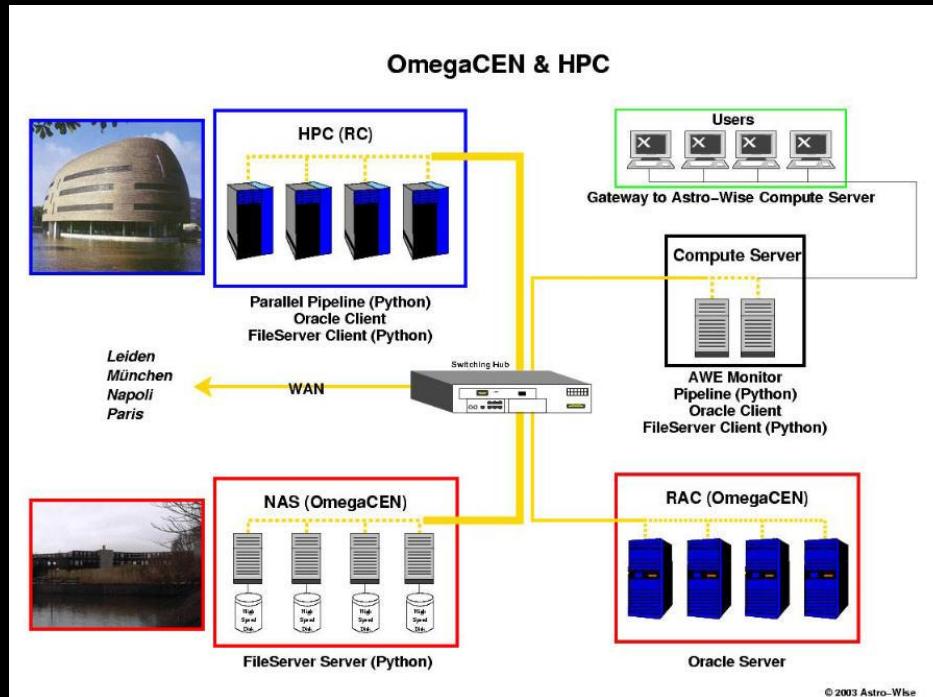


after Tyson- Szalay

The Datacentric approach

local networks and internet

Experimental
Astronomy
Vol. 35, 2013



2003 RUG-CIT

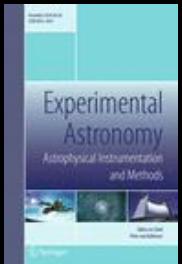
- 1-100 Tbyte – Pbytes
 - DPU- Distributed Processing Unit
 - Distributed Data server - files
 - Data manage - database - metadata
- User

AstroWise

living archives – data production



- Modeling dependencies
- 64 bit identifier
- Terabytes of pointers
- Extreme datalineage ++
- Distributed-acces-proces-calibrate- QC -analysis
- Distributed- KiDS etc
 - 11 nodes - 260 users
 - 2Pb, 45M file
 - 10 Giga table entries



Astro-WISE information system—fully datacentric

Experimental Astronomy

Vol. 35, 2013

All data beyond pixel data is Metadata

all pixel data <-> data servers

all Metadata <-> database

compute clusters / GRIDs all I/O to db

- all components scalable
- all components EU distributed

Central role db

- All I/O via db ; metadata, source catalogues
- Objects persistent in db
- data server access via global filename (key) in db
- Security
- (parallel)processing
- Webservers
- Synchronized real time National Nodes

Firefox ▾  Astro-WISE - Home  http://www.astro-wise.org/ 

Astronomical Wide-field Imaging System for Europe



a partnership of



(co-ordinator) [OmegaCEN - NOVA at Kapteyn Institute, Groningen - NL](#)
[Netherlands Research School for Astronomy, Leiden - NL](#)
[Argelander-Institut für Astronomie, Bonn - D](#)
[Osservatorio Astronomico di Capodimonte, Napoli - I](#)
[Universitäts-Sternwarte München - D](#)
[ESO, Garching bei München - D](#)
[Terapix, IAP, Paris - F](#)

An on-going project which started from a FP5 RTD programme funded by the EC Action "Enhancing Access to Research Infrastructures".

Astro-WISE Online



Overall storage and user statistics
Online storage: 1.6 PB (=1600 TB)
Number of files stored: 10717913
Database accounts: 156
Total queries¹: 514980167

¹sum for all databases since their last restart

Status of services at Astro-WISE nodes

Bonn 
Groningen 
Leiden  
München 
Napoli 

[poll details](#)

Updated: 24 Sep 2011 01:42:02

- [What is Astro-WISE?](#)
- [Guided Tour](#)
- [New user?](#)
- [Shortcuts to datasets](#)

publications
Science and technical publications

team
Organizational structure, telephone numbers etc.

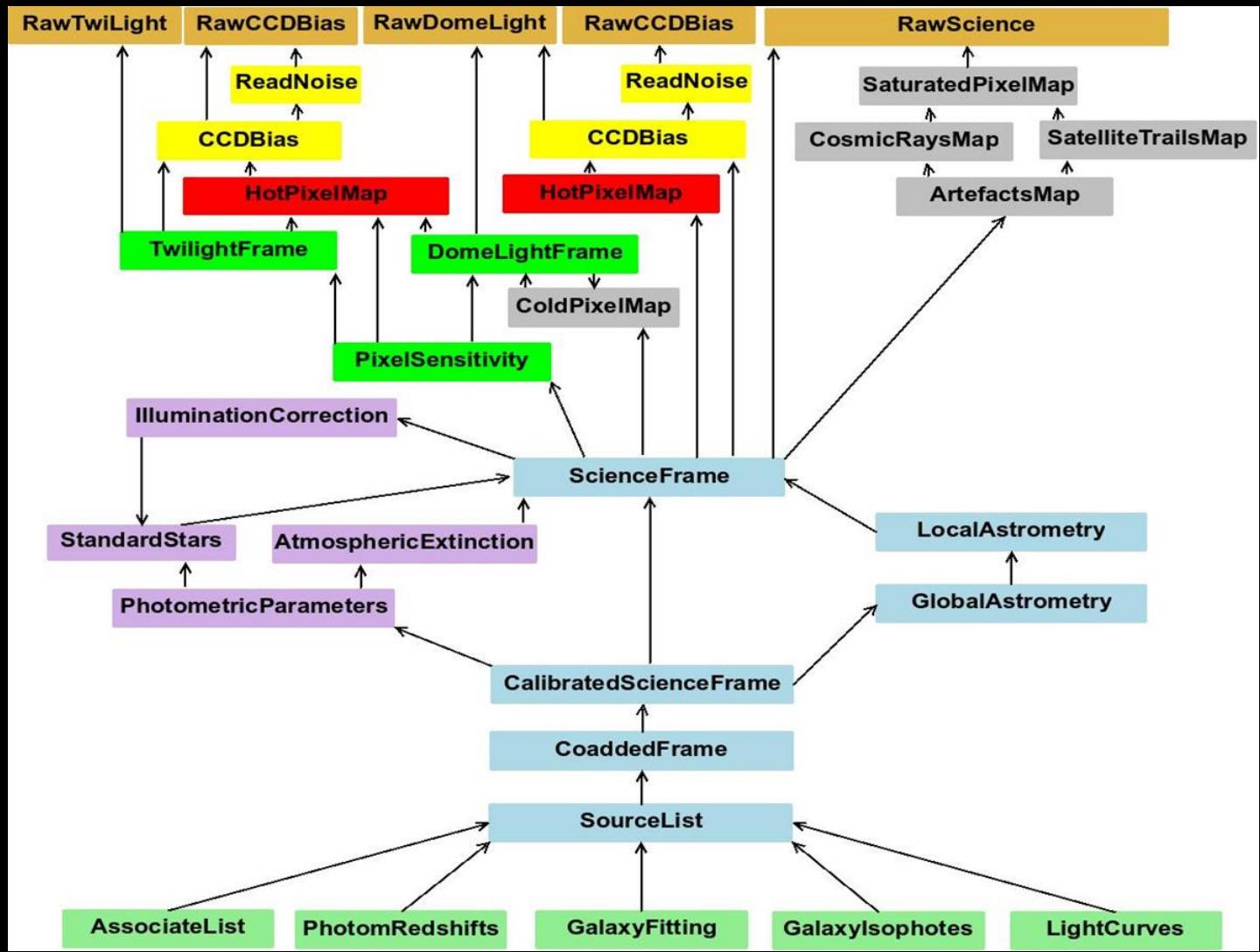
events archive
Archive of past meetings, presentations etc.

mailing lists 
[News](#)
[archive](#) [subscribe](#)
[Issues](#)
[archive](#) [subscribe](#)

web services 
[GoWise Search Engine](#)
[Database Viewer](#)
[Calibration Validation](#)
[Target Processor](#)

The universe as a spreadsheet

Target Diagram/Data lineage /backward chaining



Processing Details

creation_date	2014-05-22 18:41:01
is_valid	1
quality_flags	0
Privileges	2

Image Statistics Details

mean	+5.932e+02
median	+5.924e+02
stdev	+1.948e+01
min	+4.986e+02
max	+6.901e+02

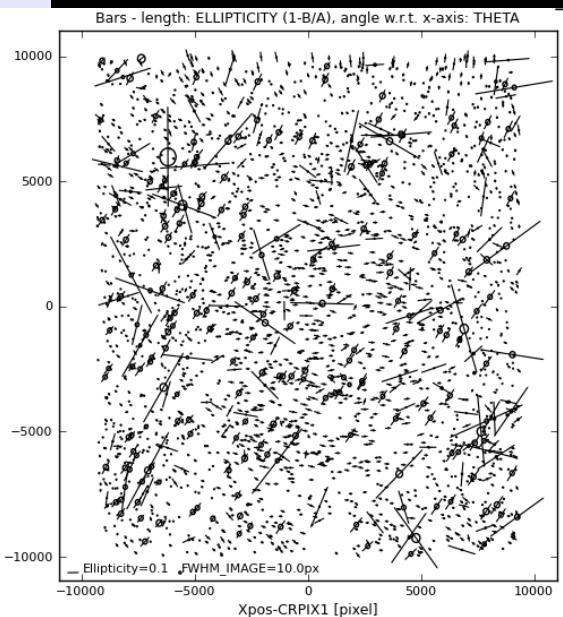
Astrometry Details
(derived from a global solution)

creation_date	2014-05-23 00:22:00
is_valid	1
quality_flags	0
SEEING	0.527
RMS	0.386
NREF	36
SIG_DRA	0.253
SIG_DDEC	0.277
MEAN_DRA	0.086
MEAN_DDEC	0.024
RMS_OVERLAP	
N_OVERLAP	
SIG_DRA_OVERLAP	
SIG_DDEC_OVERLAP	
MEAN_DRA_OVERLAP	
MEAN_DDEC_OVERLAP	

Photometry Details

creation_date	2013-10-08 16:56:48
is_valid	1
quality_flags	0
zeropoint	24.729
zp_error	0.000
zp_origin	derived
num_sources	0
extinction	0.102
ext_error	0.002

Ypos-CRPIX2 [pixel]

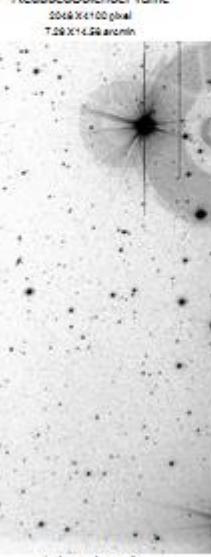


Observational Details

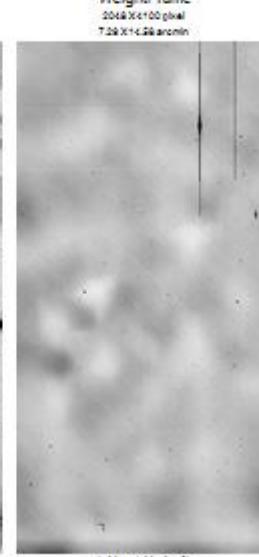
DATE_OSG	2013-05-17 04:29:06	OSERVER	UNKNOWN
OBJ_OSG	HD21910.0	ARMSTR	240.0
OBJECT	HD21910.0	ARMSTR	1.174
obs_id	T22411	ARMOND	1.165
R.A.	21:20	Fiber	OCAM_U_00-22
Dec.	0.5	mag_id	SloanR

Ch6 ESO CCD 960 of instrument
OMEGACAM

ReducedScienceFrame



WeightFrame



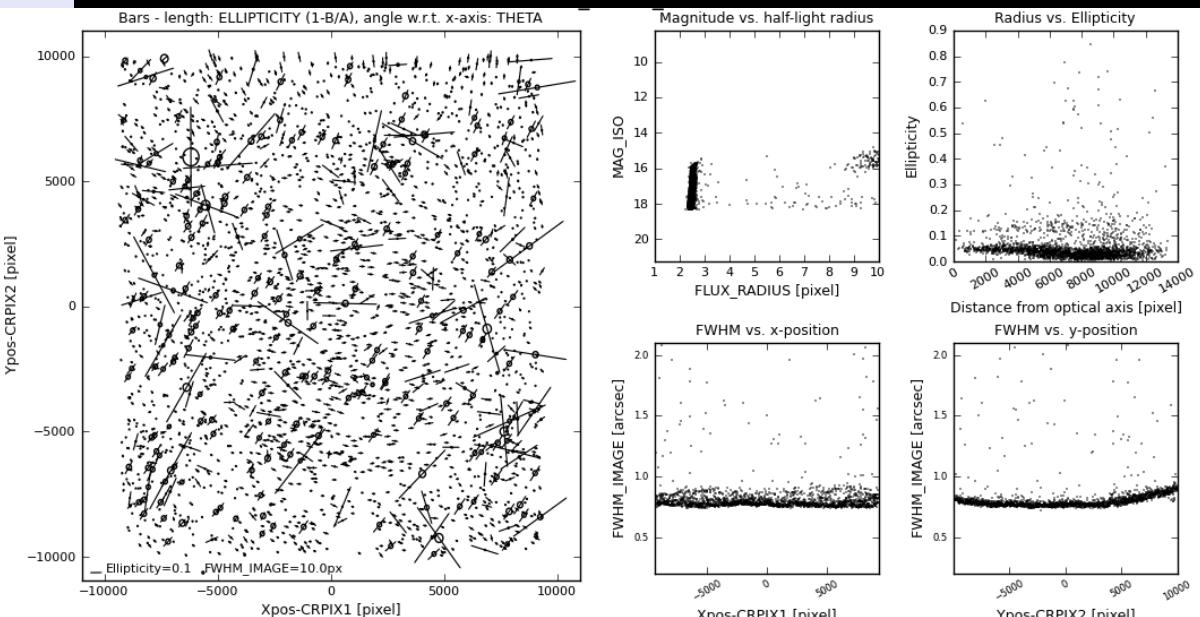
AstrometricParameters

(color map inversed)

(white = 1, black = 0)

PhotometricParameters

(white = 1, black = 0)



Re-processing view

Astro-WISE
Homepage

Target Processor

Contact
Willem-Jan Vriend

DB User
aweavalentyn

Help
Getting Started

Project

KIDS

Instrument

OMEGACAM

State

1. Preselect Target
2. Specify Target
3. Select Target(s)
4. Process or Query

Options

Preferences

Process Parameters

Upload Code

Log File
+ Messages
+ Statistics

Job overview

Query results

- A treeview is given of the target(s). This treeview gives an overview of the target's dependencies. Green dependencies mean that a newer version exists.
- For each target only one chip is shown, but with the 'view all chips' link all chips can be queried. For the processing steps, the target will be processed.
- For each target the target number, filter, date and chip are shown. With the process link on the right the target will be processed.
- On the bottom the messages, logs and statistics can be viewed.

Query Depth Calibration 2 Science -1

0	OCAM_r_SDSS SloanR	23 Feb 2014 00:06:54	ESO_CCD_#65 shown	view all chips	process	query
---	-----------------------	-------------------------	----------------------	--------------------------------	-------------------------	-----------------------

(+) (-)

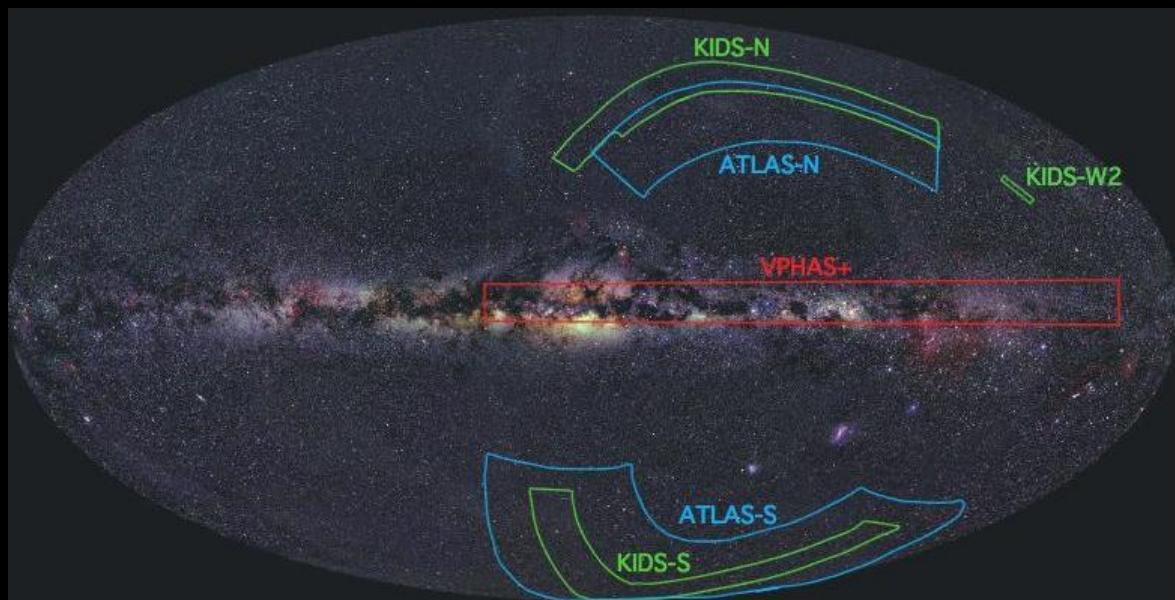
- 1.0 ReducedScienceFrame (to be built)
 - ▷ 2.1 BiasFrame
 - ▷ 2.2 ColdPixelMap (rebuild)
 - ▲ 2.3 MasterFlatFrame (rebuild)
 - ▷ 3.1 ColdPixelMap (rebuild)
 - ▷ 3.2 DomeFlatFrame (rebuild)
 - ▲ 3.3 HotPixelMap (new version)
 - ▷ 4.1 BiasFrame (new version)
 - ▷ 3.4 NightSkyFlatFrame (new version)
 - ▷ 3.5 TwilightFlatFrame (new version)
 - 2.4 FringeFrame (null)
 - ▲ 2.5 HotPixelMap
 - ▷ 3.1 BiasFrame
 - ▷ 2.6 IlluminationCorrectionFrame
 - 2.7 RawScienceFrame

Object Details	
Name	Value
type	RawScienceFrame
filename	OMEGACAM.2014-02-23T00:06:54.704_1.fits
chip	ESO_CCD_#65
filter	OCAM_r_SDSS
Object	STD,ZEROPOINT
quality_flags	0
is_valid	1
observation	23 Feb 2014 00:06:54
creation	23 Feb 2014 06:38:38
Links	Download Preview DBview Quality

ESO public surveys

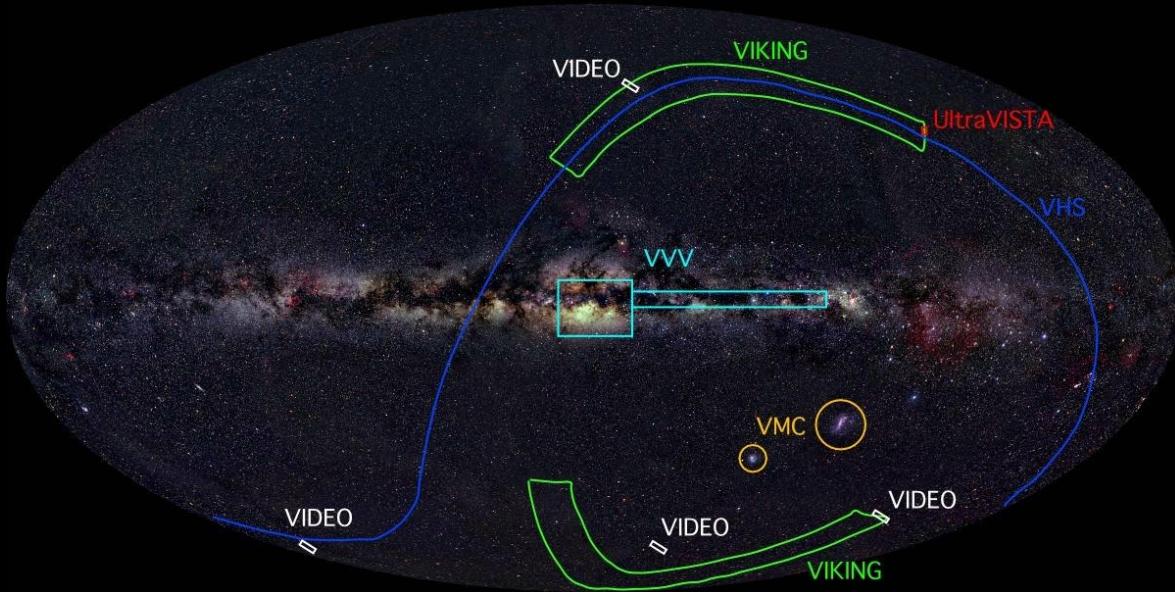
VST:

- ▶ ATLAS
- ▶ VPHAS+
- ▶ KiDS



VISTA:

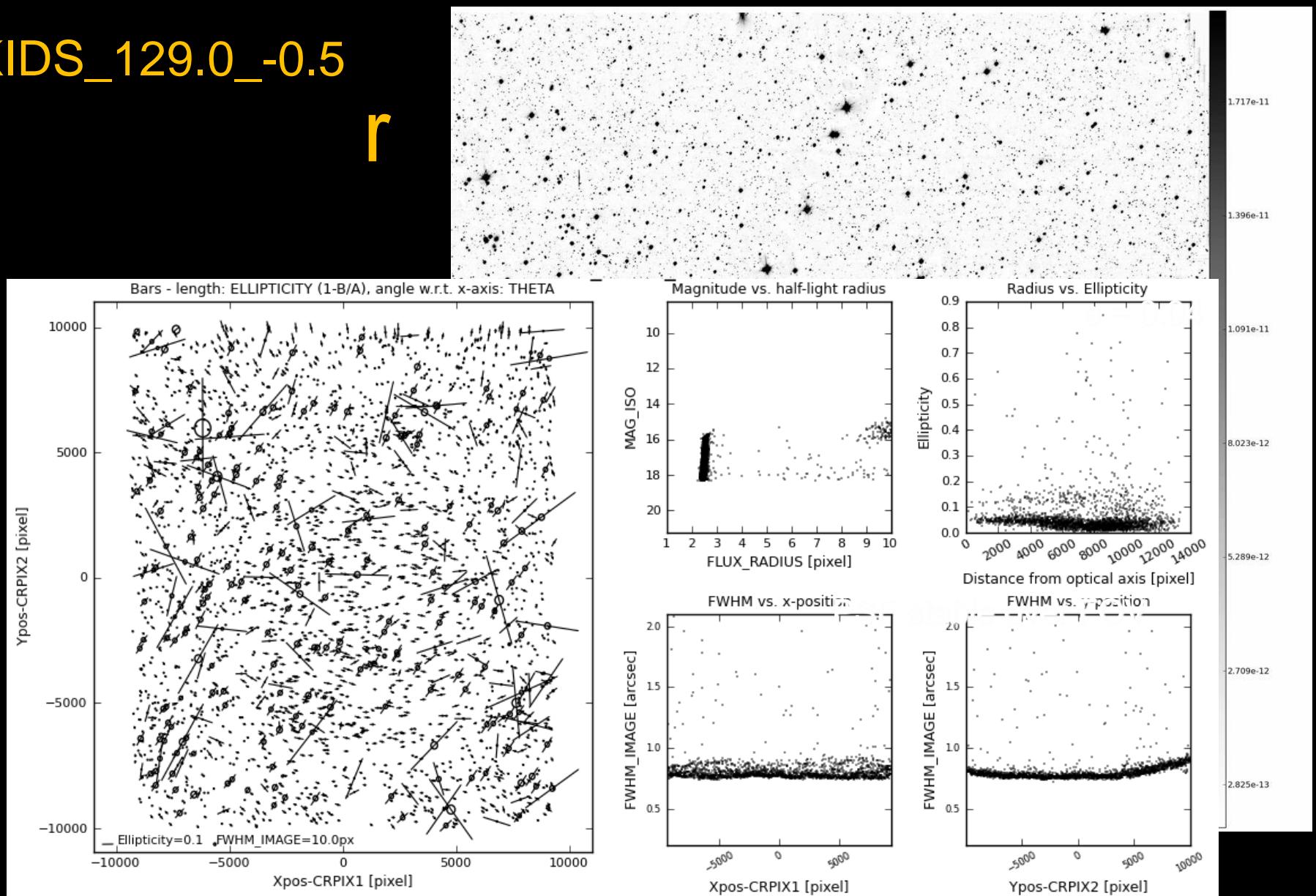
- ▶ VHS
- ▶ VIKING
- ▶ VIDEO
- ▶ UltraVISTA
- ▶ VIDEO
- ▶ VVV



KiDS: data quality

KIDS_129.0_-0.5

r



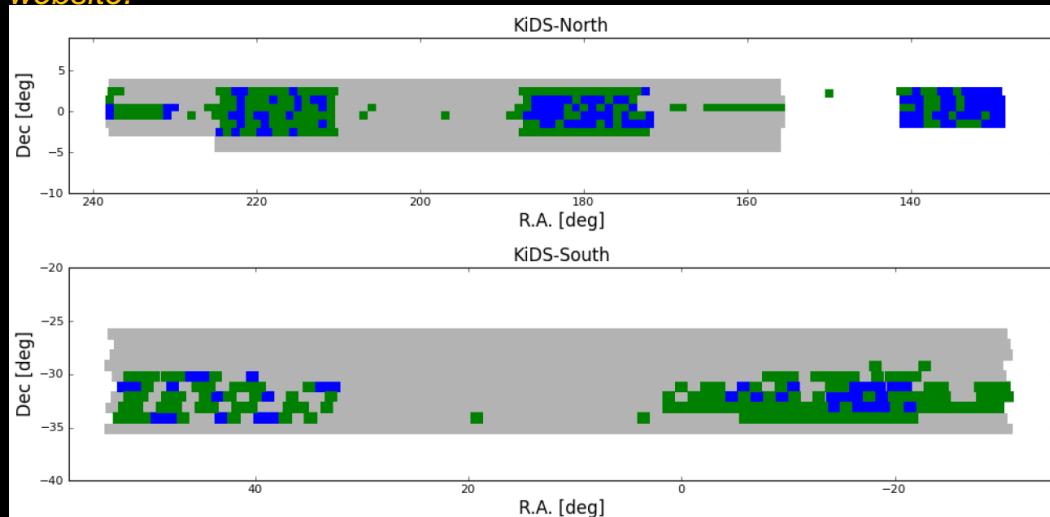
KiDS data releases produced by Astro-WISE

3 Public Releases, via ESO

		#tiles	#sq.deg.	#sources	#files	#Tbyte	Reference
ESO-DR1	completed tiles < 1 Oct 2012	50	56	5.7E6	800	0.7	de Jong+15 A&A 582,62
ESO-DR2	completed tiles < 1 oct13	98	110	1.1E7	1568	1.4	de Jong+15 A&A 582,62
ESO-DR3	completed tiles \lesssim 1Oct15	292	328	3.3E7	4672	3.6	de Jong+16, in prep
TOTAL	Oct11-Oct15	440	495	50 M	7040	5.7	

Plus 5 Internal Releases, KiDS consortium-only

Released through **3 portals**: Astro-WISE, ESO archive and KiDS website.



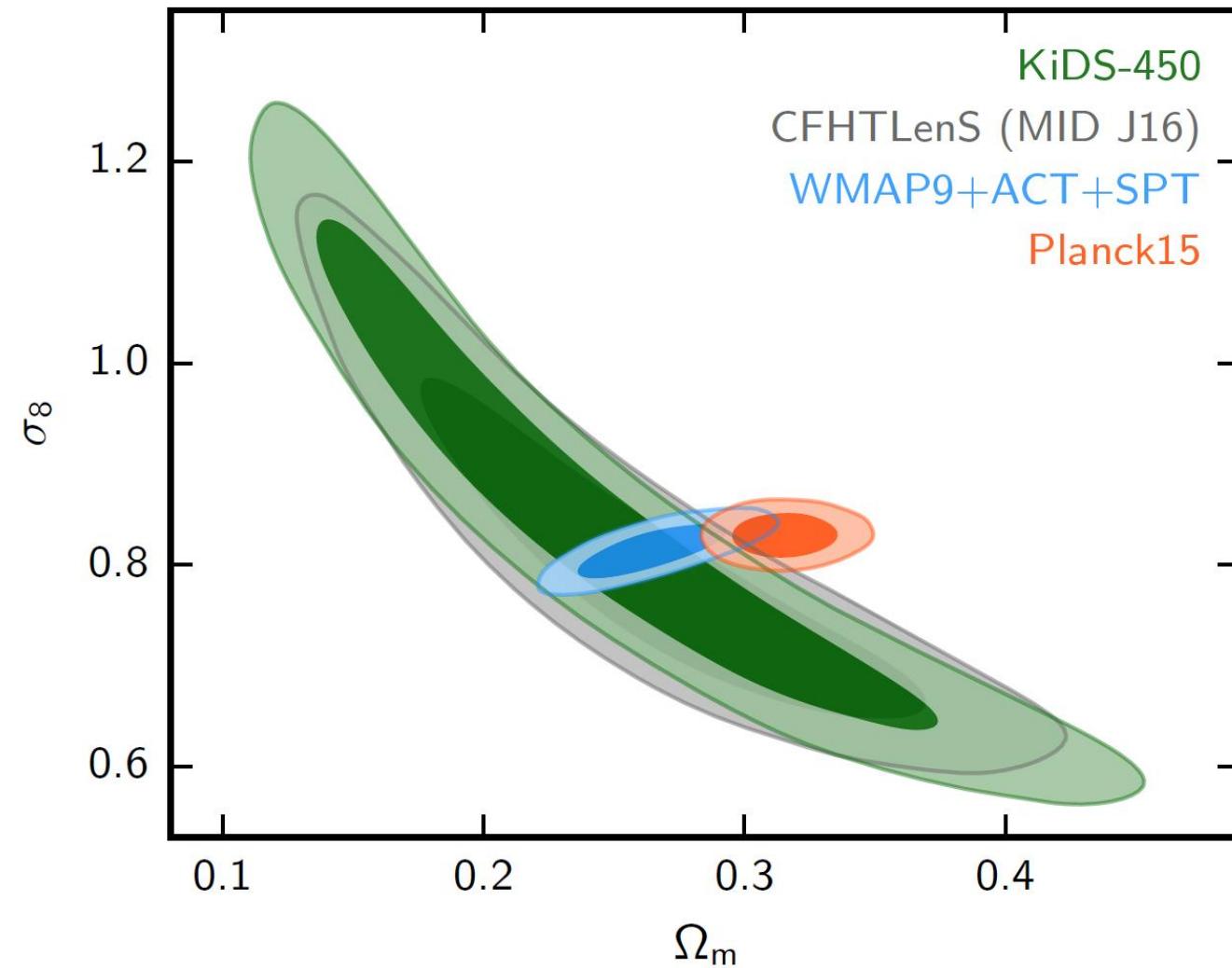
Survey products public releases

- single exposures + coadds
- weights, auto + manual masks
- single- and multi-band catalogs
- aperture and PSF homogenized
- basic star/galaxy parameters
- photometric redshifts
- quality parameters!

Consortium internally has released more auxiliary products: KiDS+VIKING 9-band photometry, galaxy morphometry, group catalogs, shear maps, ...

KiDS

KiDS-450: Result



TarGet

98 99 100 101 102 103 104



MuseWise

ESO VLT
IFS 24 IFU
100 Gb/night

[Home](#) | [Contact](#) | [Help](#) | user awjbrinchmann | project INM | Preferences | Tables

Welcome to the MuseWise DBView Web Service

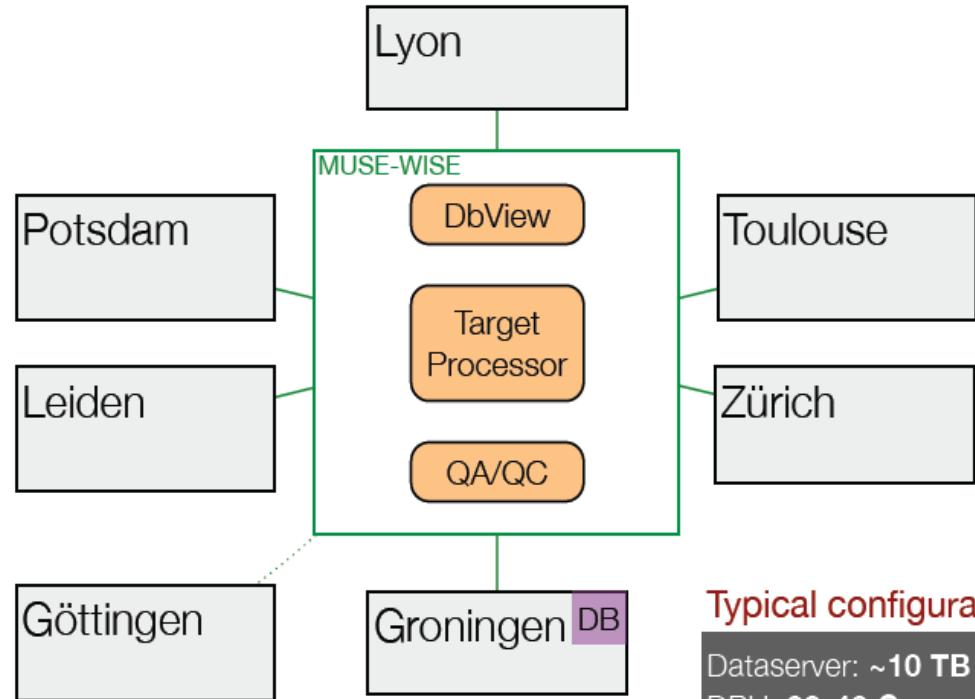
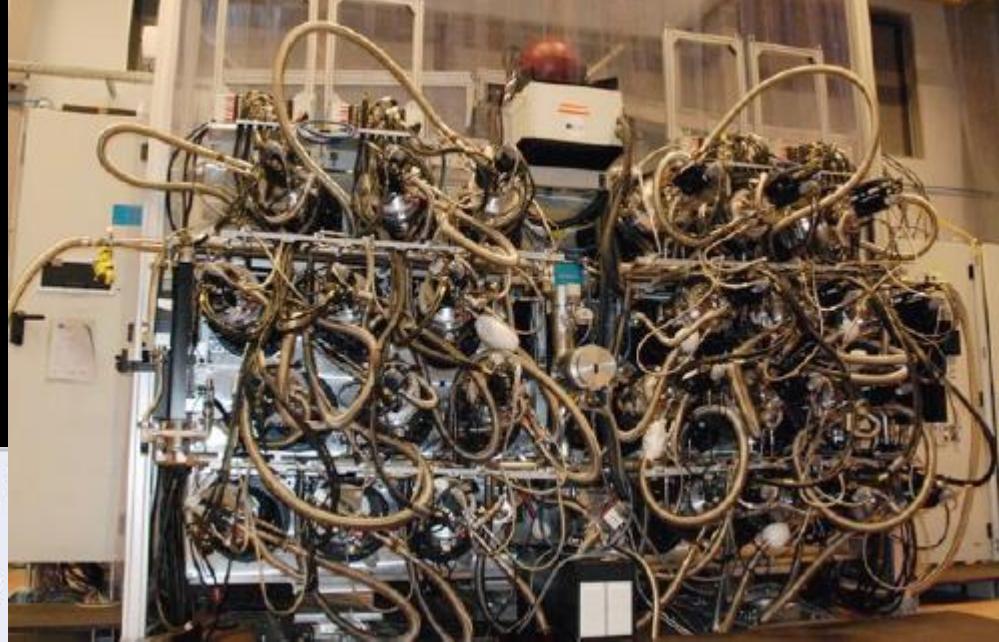
The following table lists the versions of MuseWise components

Component	Version
Muse-WISE version	0.03.01
muse2wise version	1.58
QC version	v0.0.1
musep version	0.06.00
CPL version	0.5

Please choose a table category to start querying

- All tables
- Raw Frames
- External Products
- Processed Calibration Products
- Processed Science Products

empowered by



From DRS to WISE

```
XML  
<xml>  
<frame tag="MASTER_BIAS">  
<parameter name="Observer">
```

Generate Python from XML

Python

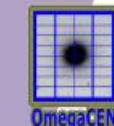
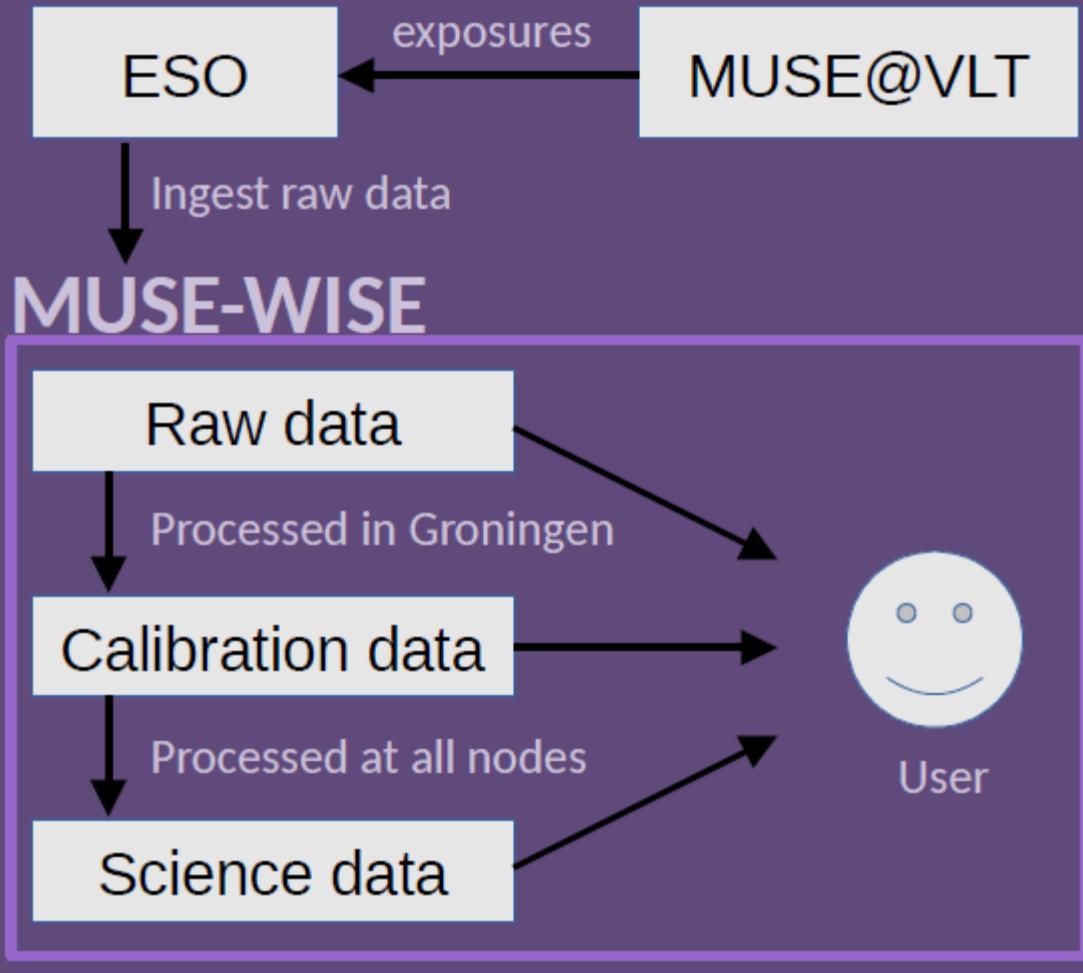
```
Class MASTER_BIAS(DataObject):  
  
    Observer = persistent(str)
```

Generate Database from Python

Database

```
TABLE MASTER_BIAS  
Column ID  
Column Observer  
Column Filename
```

Data-flow





MUSE
1' x 1'
HDF-S

Ly- α

Wisotzki
et al
2016

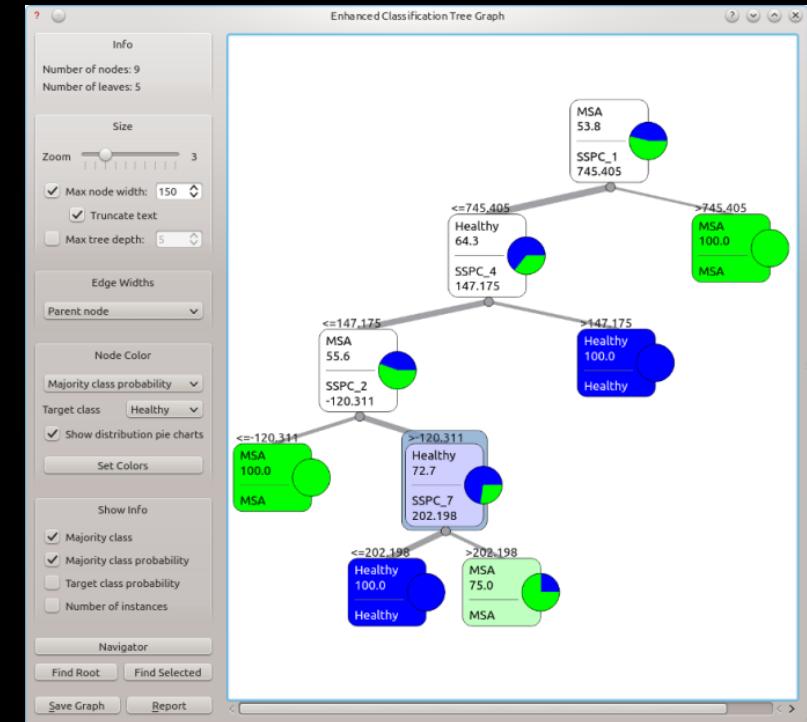
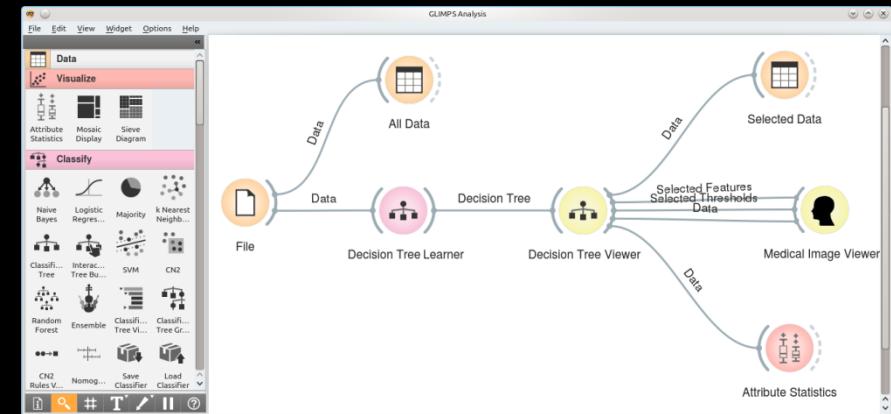
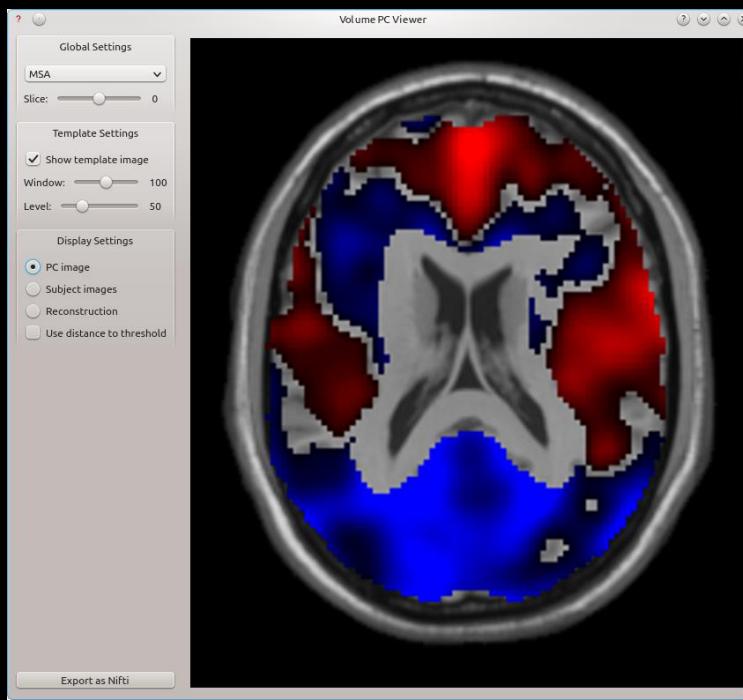
TarGet GLIMPS - Visualization of decision trees

Orange -Interactive

Pet scans

Data lineage

SAMP



National Archive 1000 M files

The Reading Monk - Mozilla Firefox

Bestand Bezoeken Beeld Geschiedenis Bladwijzers Extra Help

http://rugtest5.service.rug.nl:9001/Monk

Webservices < Target < Rijksunivers... > Astro-WISE Processing > Book flights > The Reading Monk

MONK



zoeken

There are 152 results. Showing results 1 - 10 for manuscript

[show paragraph](#) [show page](#) [Query image](#)

onderwijzer J.G. Blieën te Amsterdam
— Bestelt fiet
Vervolg op blad 64.

onderwijzer J.G. Blieën te Amsterdam — Bestelt fiet Vervolg op blad 64.

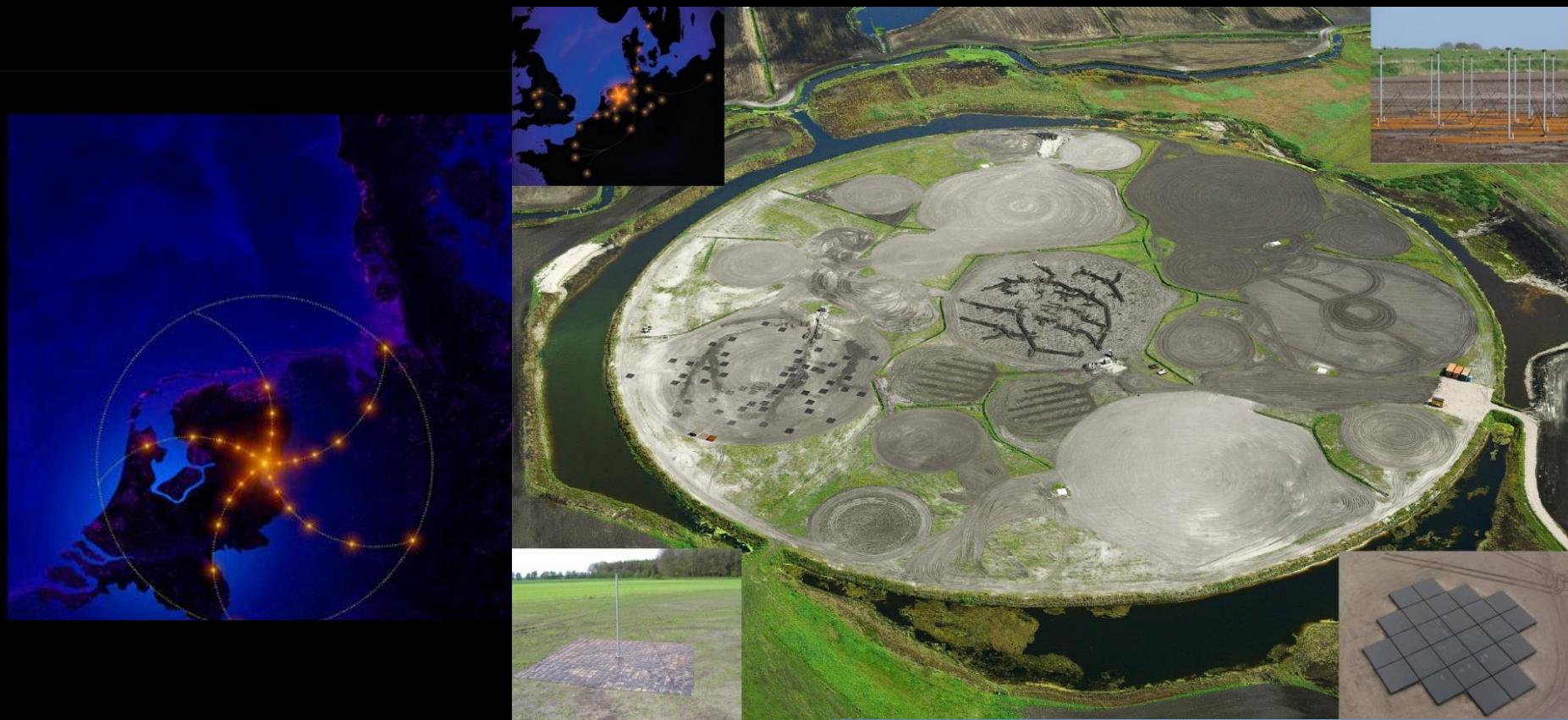
[show paragraph](#) [show page](#) [Query image](#)

colonial Etablissement te Amsterdam
Dua Misla... Yaa Doo...
colonial Etablissement te Amsterdam

[show paragraph](#) [show page](#) [Query image](#)

Hoff Seeyver te Amsterdam tot het keuren van

LoWISE – LOFAR LTA



5 Petabyte/year

SARA-Amsterdam

Julich

EoR -lineage

Lofar Archive Mozilla Firefox

File Edit View History Bookmarks Tools Help
<http://lofar.astro-wise.org/Lofar/>?latest=1&product>All&submitted=1

Most Visited openSUSE Getting Started Latest Headlines Mozilla Firefox

Lofar Archive

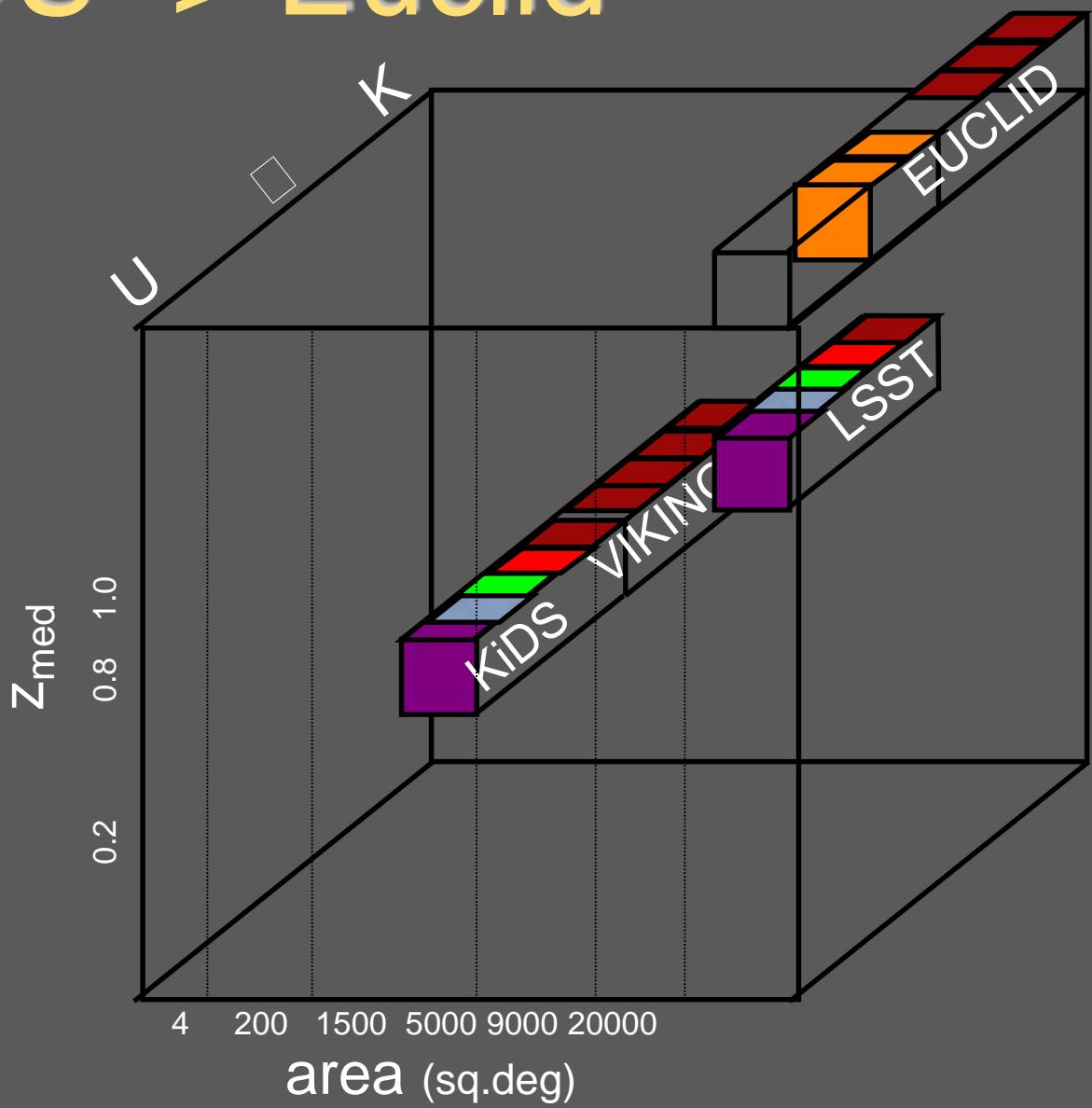
16 results for UVMeasurement

edit columns

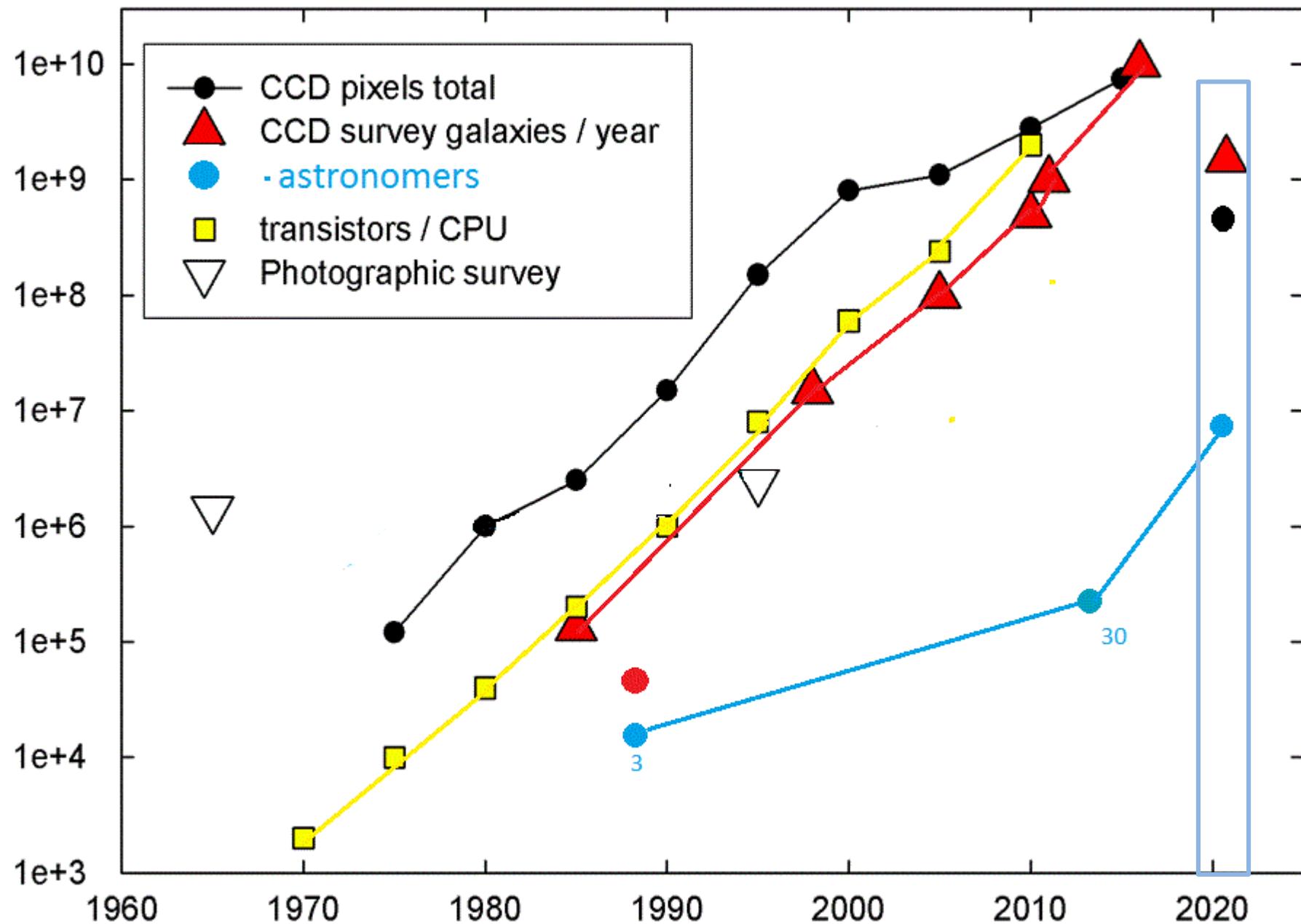
Project	antennaSelection	CreationDate	InstrumentFilter	measurementIdentifier	measurementType	startTime
1 test-lofar	HBA Both	2010-04-06 12:28:03	170-230 MHz	10142	Target	2010-02-06 12:01:00
2 test-lofar	HBA Both	2010-04-06 10:26:43	170-230 MHz	10159	Target	2010-02-06 12:01:00
3 test-lofar	HBA Both	2010-04-02 22:59:47	170-230 MHz	10015	Target	2010-02-06 12:01:00
4 test-lofar	HBA Both	2010-04-02 22:51:32	170-230 MHz	10016	Target	2010-02-06 12:01:00
5 test-lofar	HBA Both	2010-04-02 22:22:51:63	170-230 MHz	10018	Target	2010-02-06 12:01:00
6 test-lofar	HBA Both	2010-04-02 22:22:51:63	170-230 MHz	10027	Target	2010-02-06 12:01:00
7 test-lofar	HBA Both	2010-03-31 13:23:00	170-230 MHz	10014	Target	2010-02-06 12:01:00
8 test-lofar	LBA Sparse Even	2010-03-15 16:06:41	10-90 MHz	10009	Target	2010-03-06 12:01:00
9 test-lofar	LBA Sparse Even	2010-03-12 16:57:27	10-90 MHz	31	Calibration	2010-03-09 15:55:01

LOFAR

KiDS \rightarrow Euclid



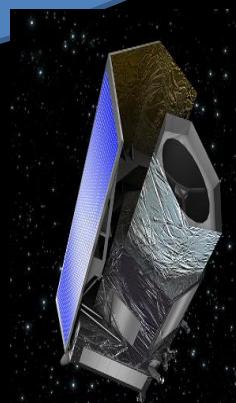
Trends in Optical Astronomy Survey Data



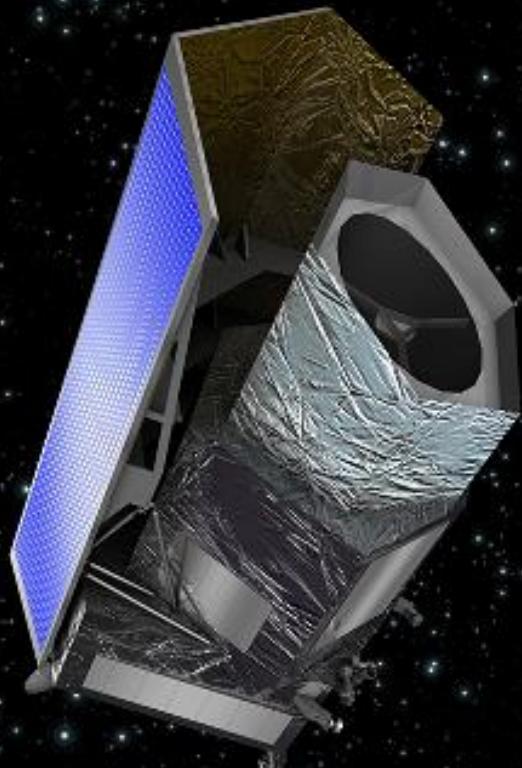
Distributed communities acces-proces-calibrate-analyse publish

Euclid:

- 1300 registered members and growing
- 800 researchers
- 130 laboratories/departments
- 13 European countries contributing: Austria, Denmark, France, Finland, Germany, Italy, Netherlands, Norway, Spain, Switzerland, Portugal, Romania, UK + USA IPAC
- NASA/US: provides the IR detectors.



Euclid



ESA

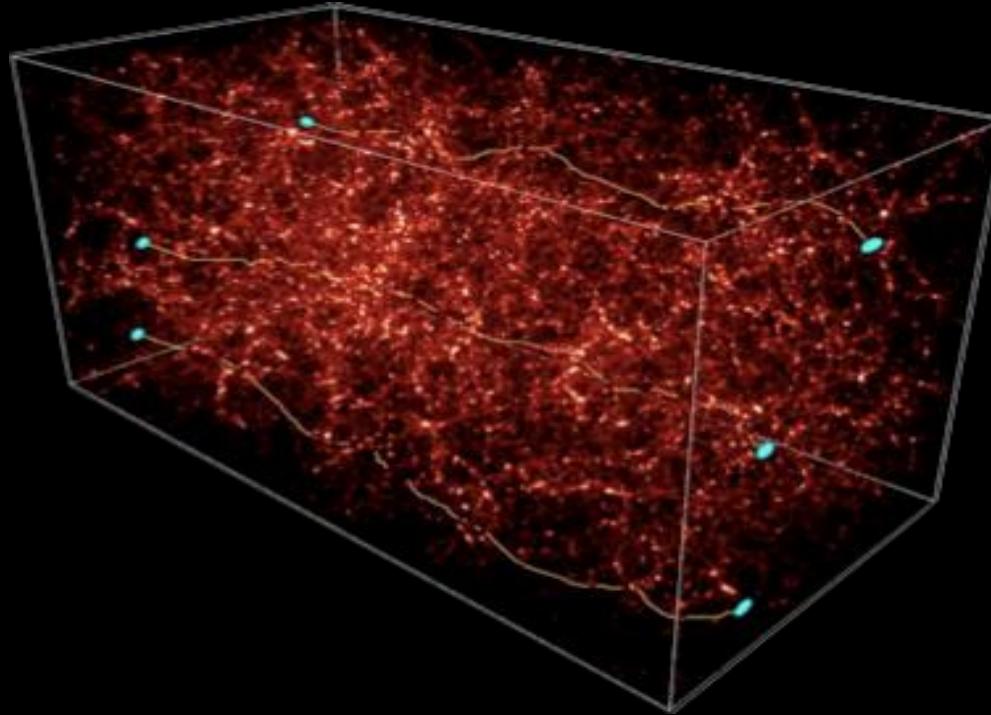
Due for launch in 2020

Euclid Archive System (EAS) is
data centric information system
with many of the WISE concepts

EAS prototype uses Astro-WISE
code to provide a functional system
for development

EAS prototype hosted in the Euclid
SDC-NL in Groningen

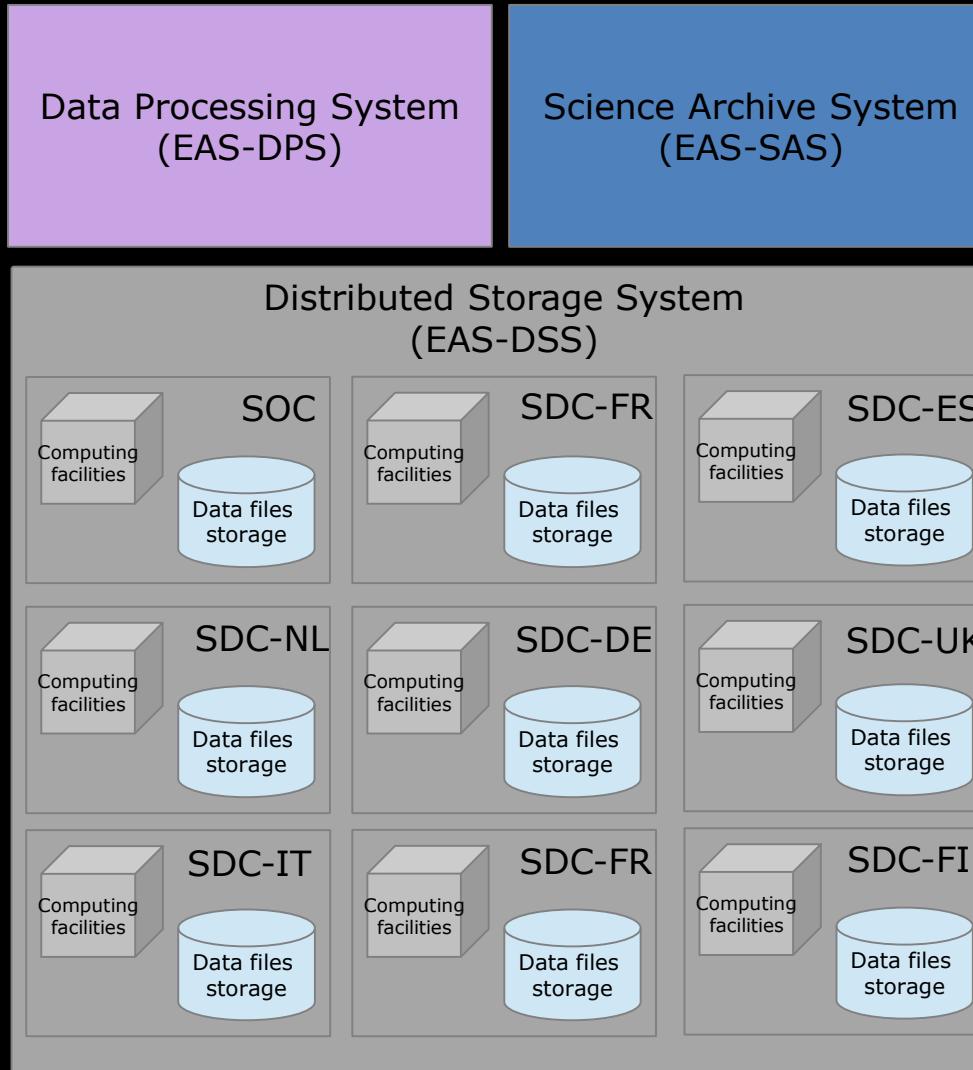
Weak gravitational lensing as probe of dark matter

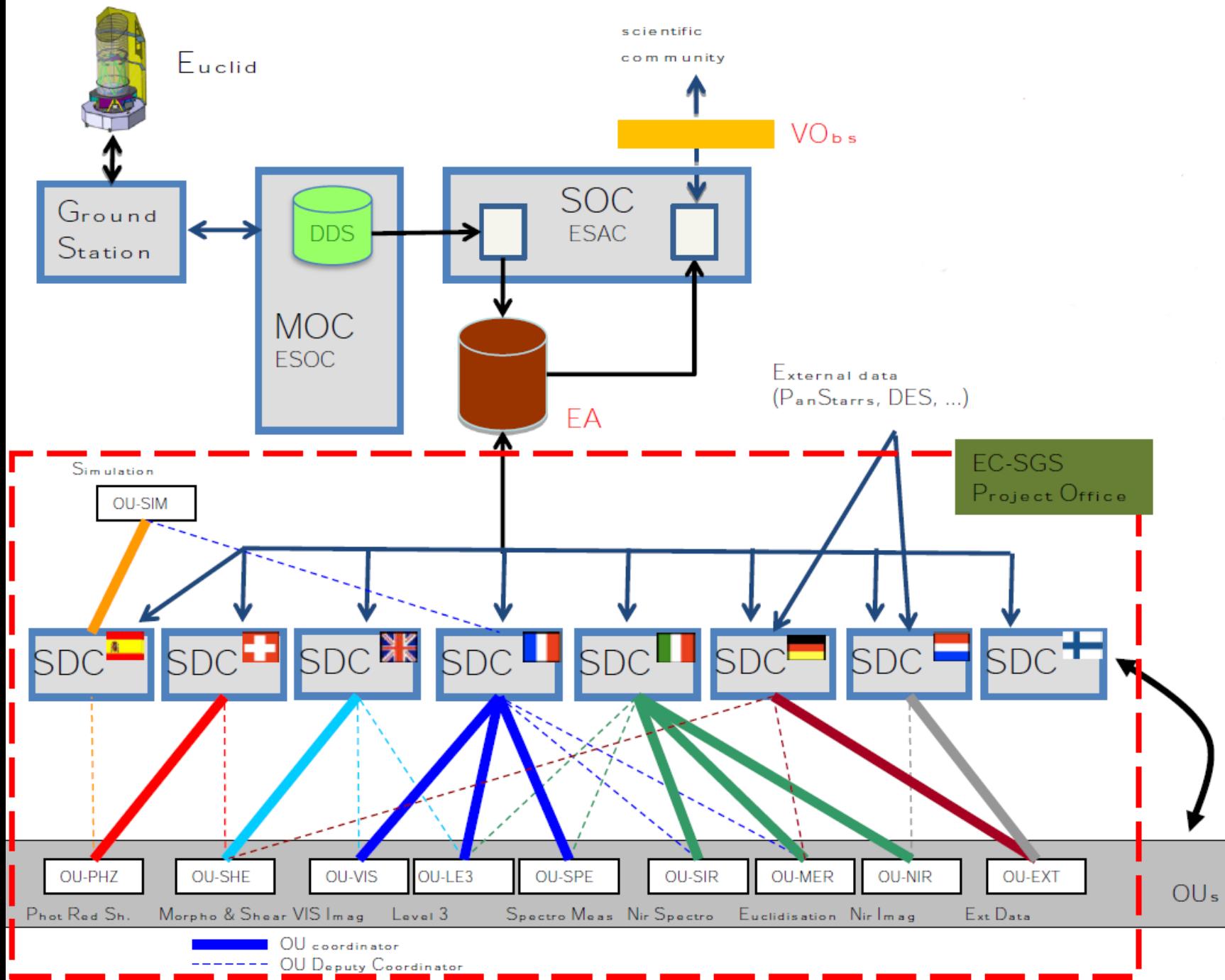


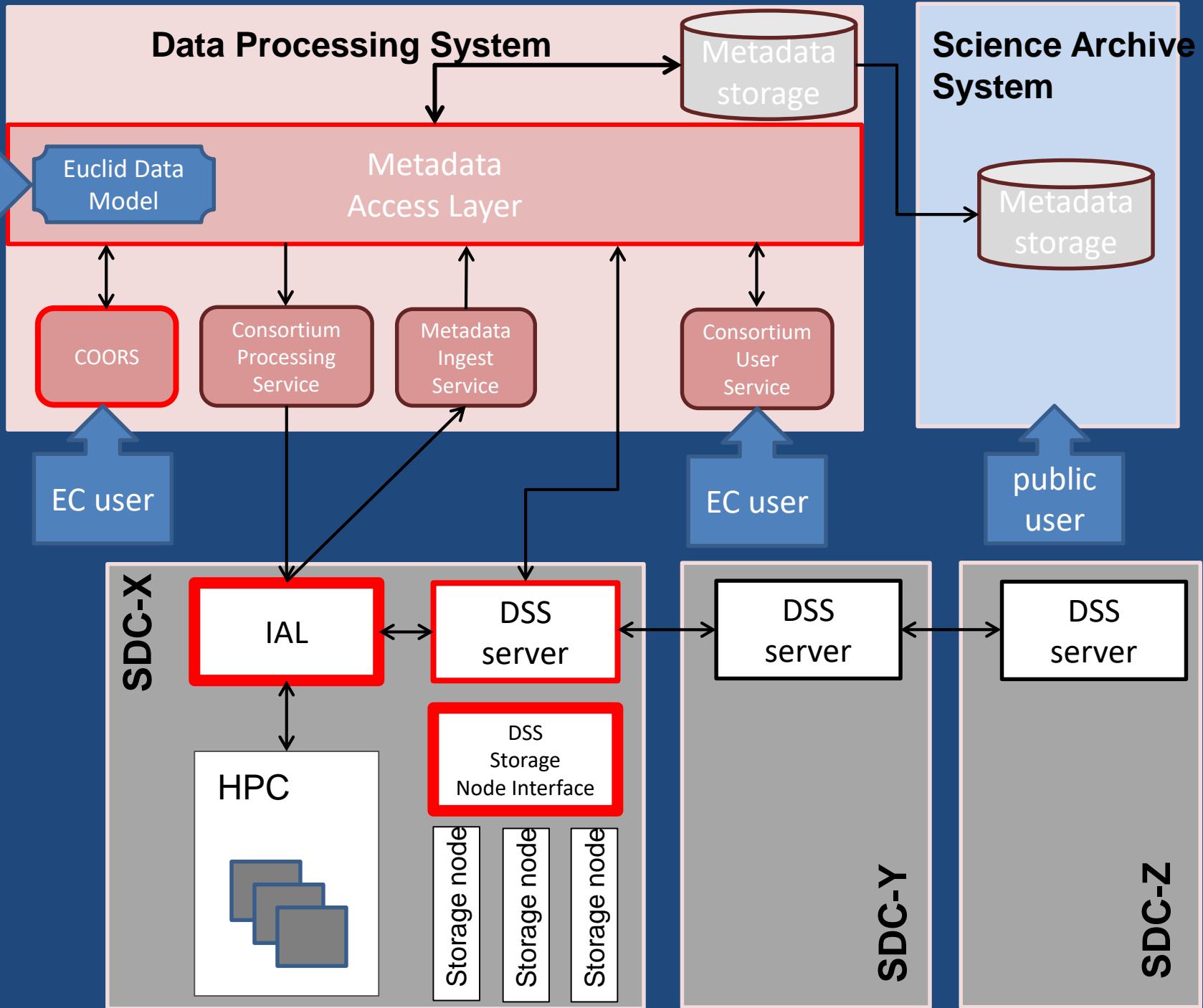
KiDS: $< 100 \ 10^6$ redshifts
EUCLID: $1.5 \ 10^9$ redshifts

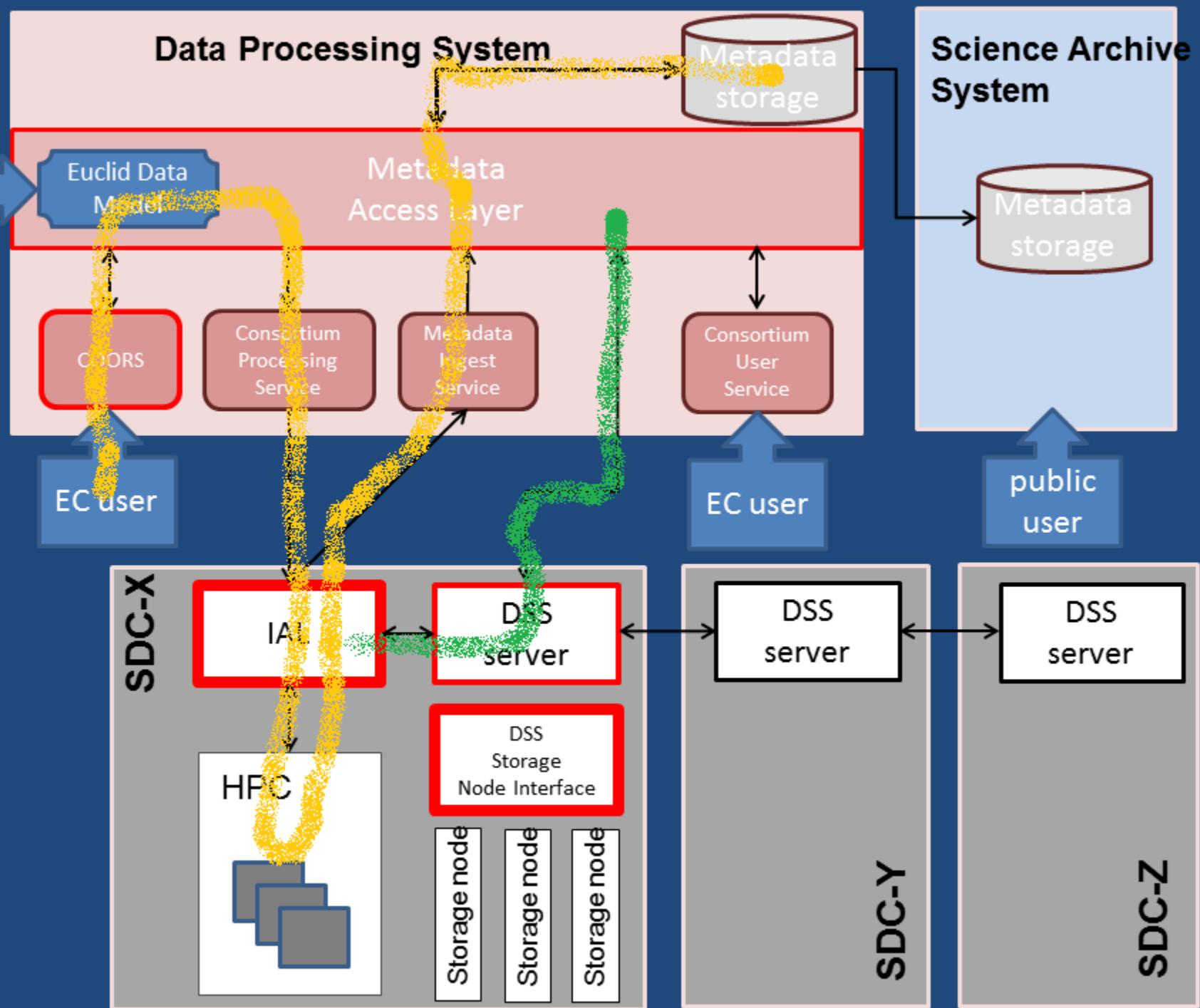
Every galaxy has its own 4 PSFs
QC- bias – re-processing

EAS Design









Data federations

= Focus on projects – disciplines – domains

- Standards -Project management – Sociology – Security
- Acces layers
 - zoo of hardware / hardware operations
- datacentric
- full reproducibility – lineage
- QC – re-processing
- Conversion from Data Model to database schema
 - Astro-WISE - Python-classes
 - MUSE - XML
 - Euclid - XSD -classes- XML-objects

