Virtual Observations and Data Mining in Astronomy

Future directions in Astronomical data management

Kapteyn Astronomical Institute
University of Groningen

Edwin A. Valentijn

11 Okt 2010

E-science

- Beyond “workstation science” of the 80-90’s
- Distributed services
- Distributed communities
- Distributed archives
- p2p networks – KAZAA- NAPSTAR
  - Share cpu
  - Share storage
  - Share info / meta data /knowledge

future

- Lofar 2010 www.lofar.org
- VST-OMEGACAM 2011 www.astro.rug.nl/~omegacam
- GAIA 2012
- ALMA 2013 www.eso.org/sci/facilities/alma
- EUCLID 2018
- LSST 30Tb/night www.lsst.org
- SKA 2022 www.skatelescope.org

Basics- Surveys

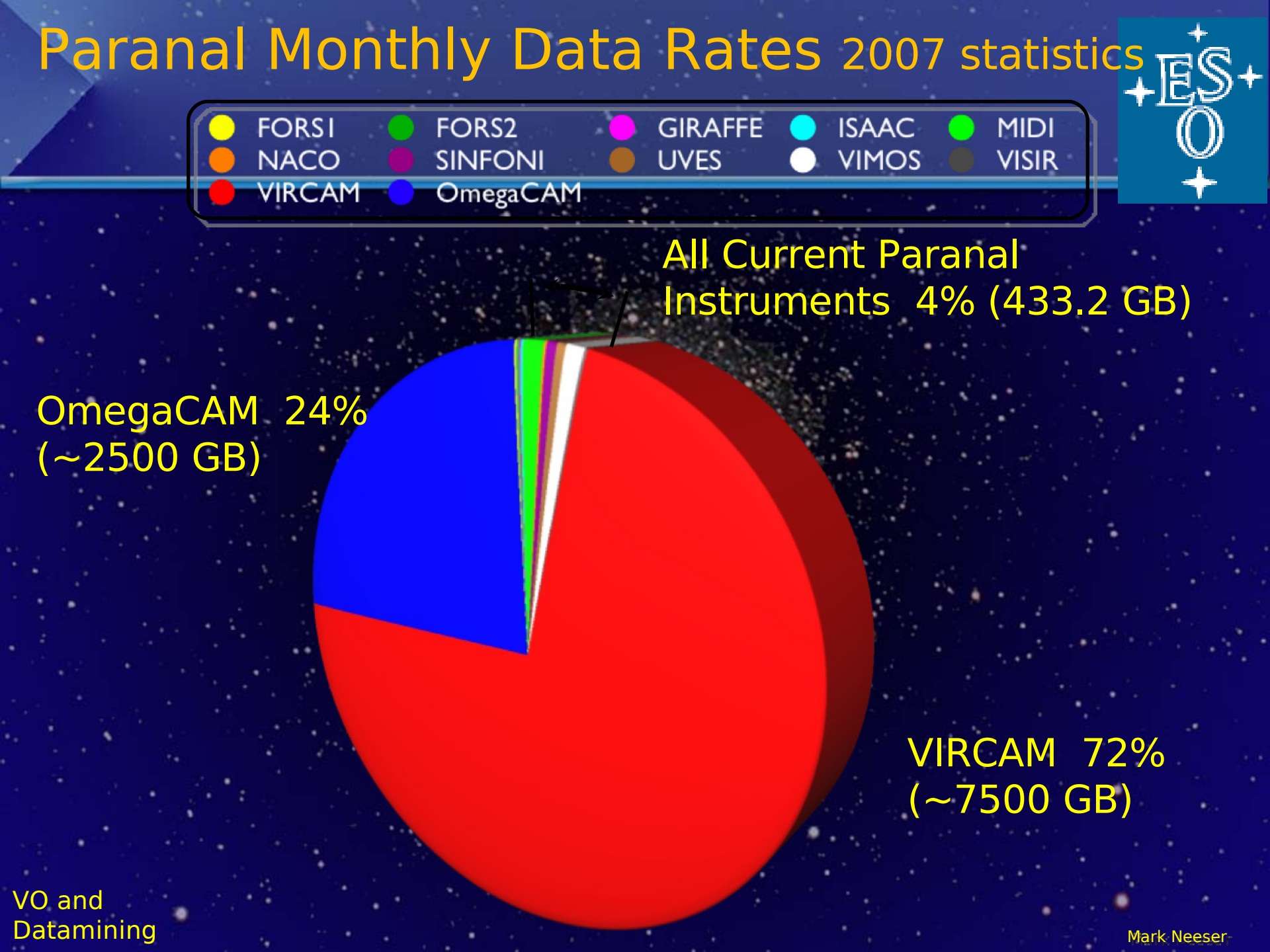
- Defined area on sky
- Homogeneous
  - Survey limit
    - Flux (magnitude)
    - Size
    - Surface brightness
    - distance
- Quality control

Basics - time

Everything changes in time

- Physical changes
- Our inside in modeling
- Methods, code, bugs

Paranal Monthly Data Rates 2007 statistics



Astro-WISE information system – fully datacentric

All data beyond pixel data is Metadata

all pixel data <–>data servers
all Metadata <–> database

compute clusters / GRIDs all I/O to db

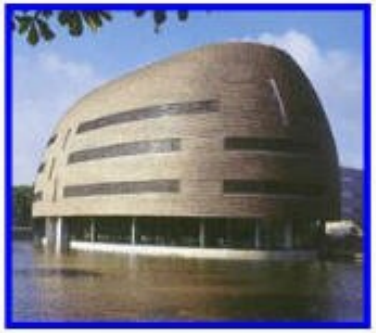
- all components scalable
- all components EU distributed

N params N data back to basics

- Joins – links
- ++ Inheritance – dependencies

- Everything in cs is adresses
  memory, ASCII, namespaces, registry
- Optimize , organize, index
- management

VST - Virtual Survey Telescope



HPC (RC)

Parallel Pipeline (Python)
Oracle Client
FileServer Client (Python)

Users

Gateway to Astro-Wise Compute Server

Compute Server

AWE Monitor
Pipeline (Python)
Oracle Client
FileServer Client (Python)

Leiden
München
Napoli
Paris

WAN

Switching Hub

NAS (OmegaCEN)

FileServer Server (Python)

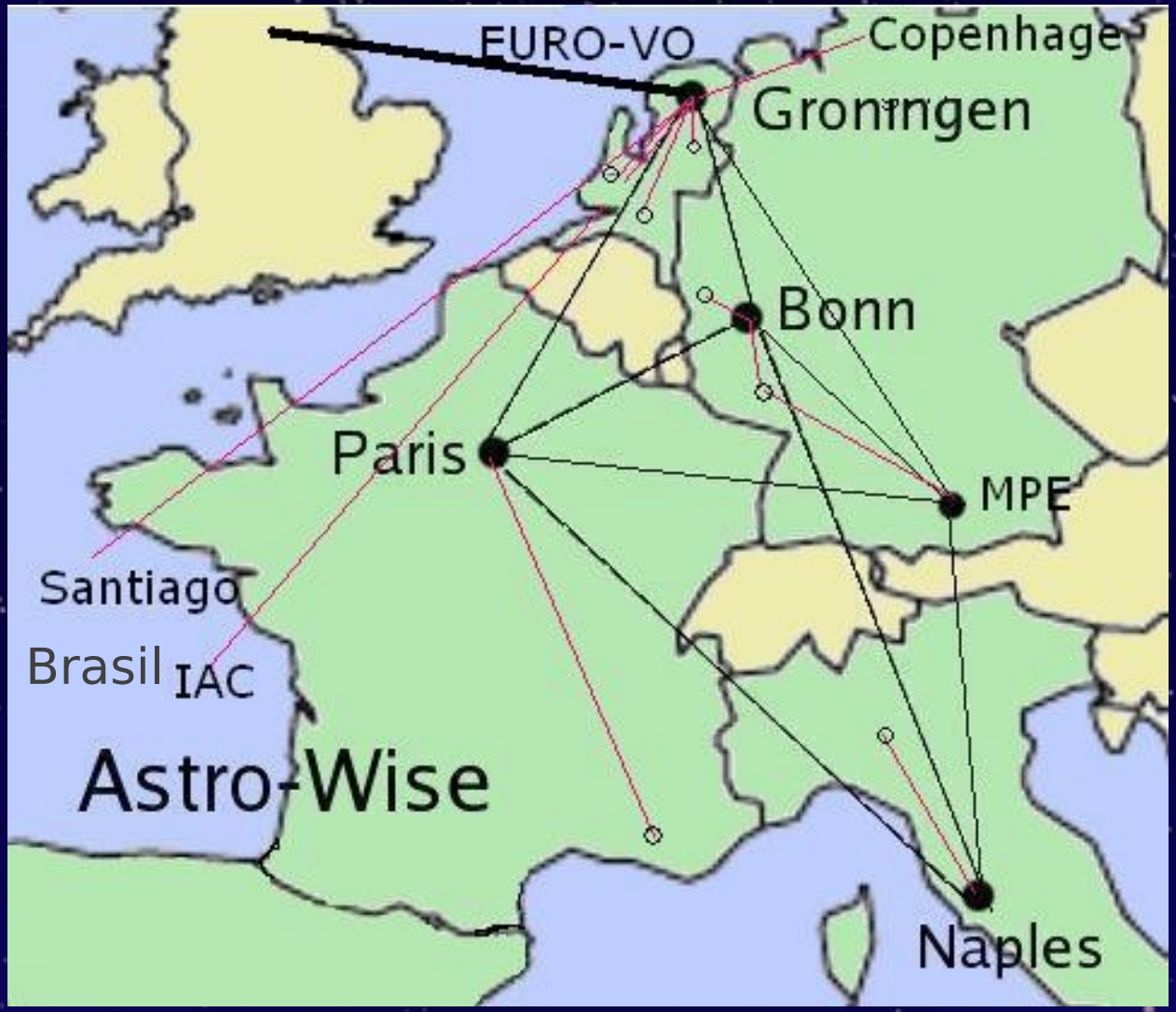
RAC (OmegaCEN)

Oracle Server

VO and
Datamining

© 2003 Astro-Wise

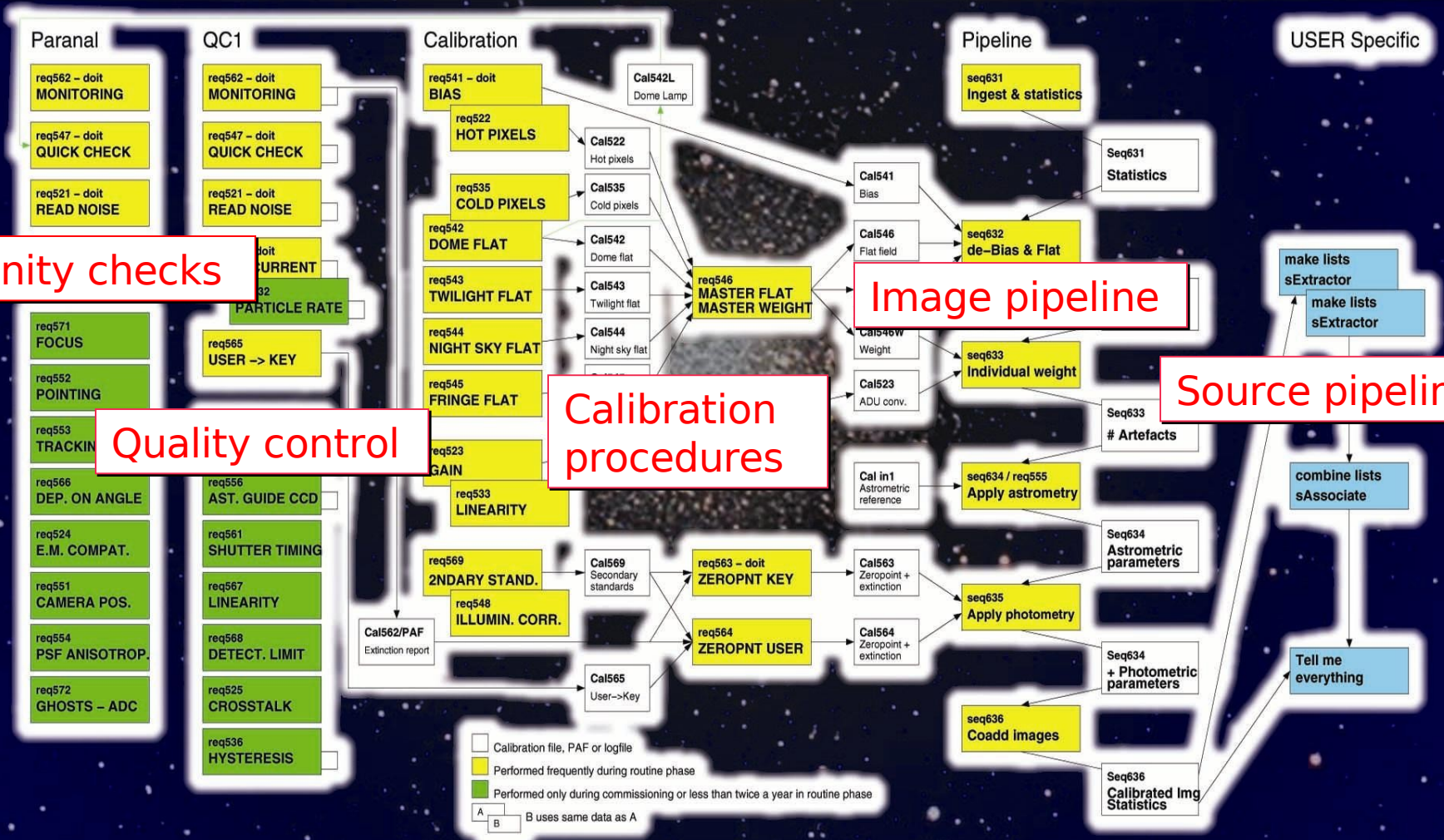
Centers – satellites



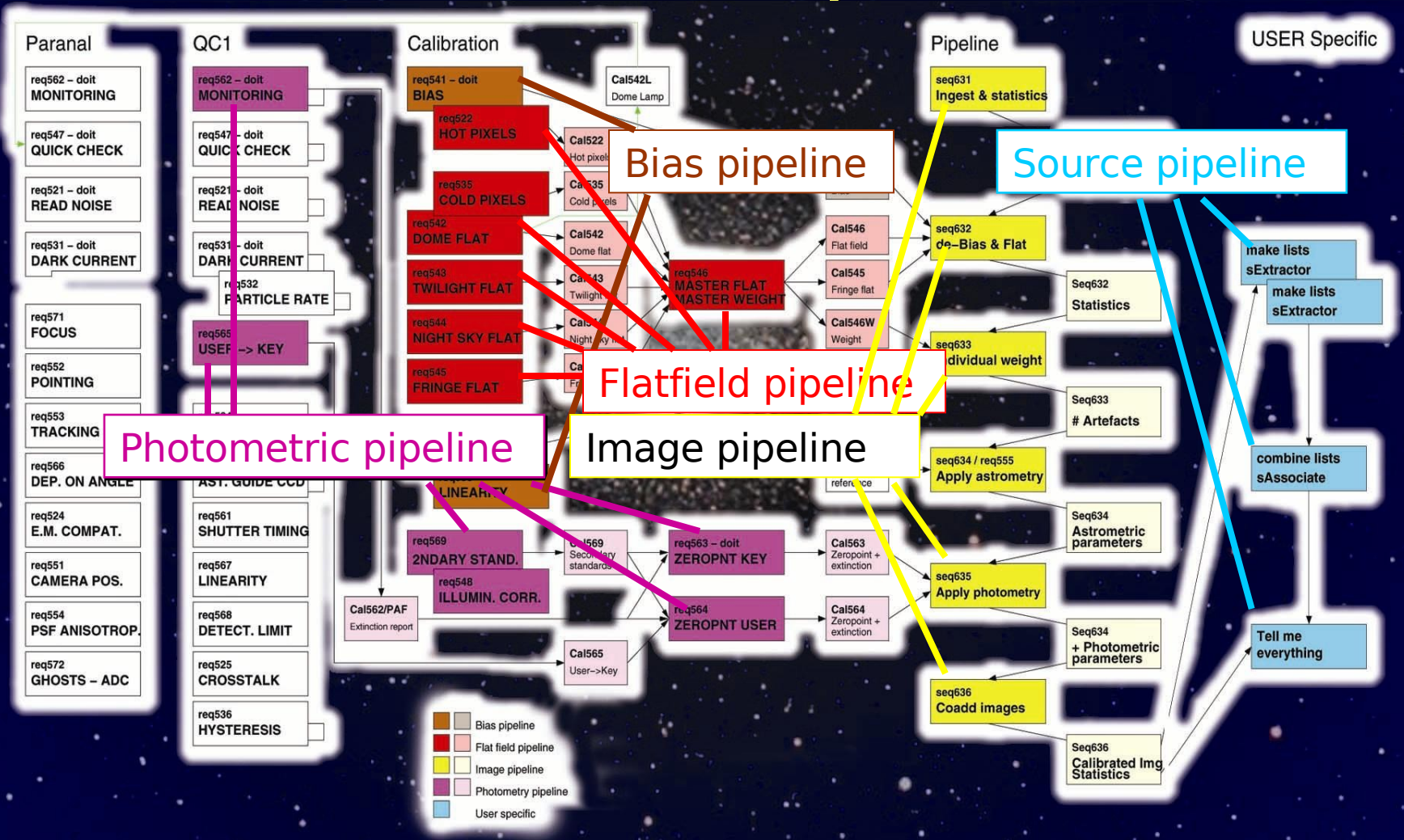
basics- pipelines

- Workflow
- What triggers a pipeline?
  - Data items
  - Operators
  - users

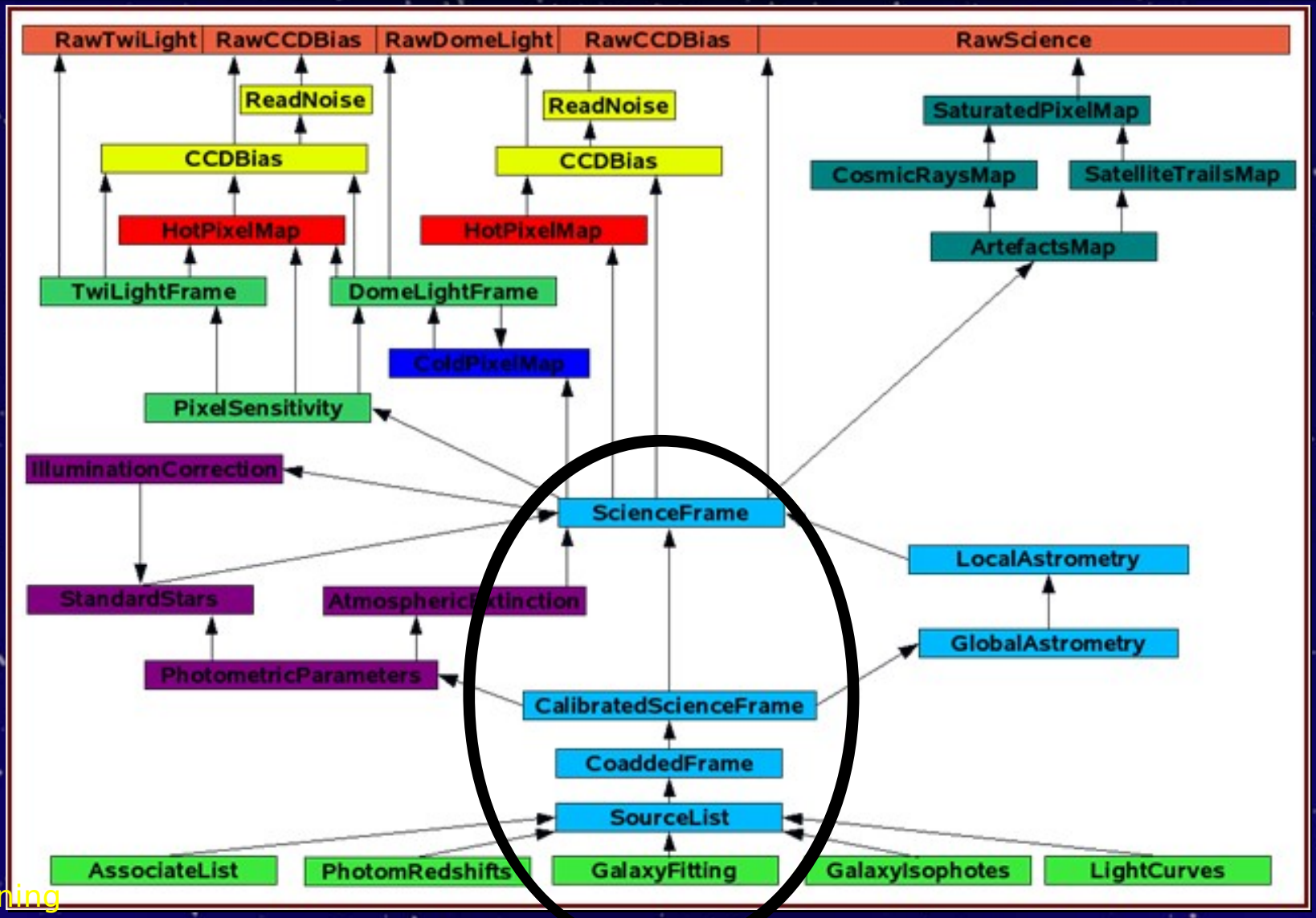
basics- Data Model / flow



Basics
Astro-Wise Pipelines



TARGET diagram



Target processing: ++ the make metaphor

```
awe> targethot=HotPixelMap.get(date='2003-02-14', chip='A5382')
```

The processing chain is

ReadNoise <-- Bias <-- HotPixels

- > class HotPixelMap(ProcessTarget):
- > > def self.make()

- > class ProcessTarget():
- > > def get(date, chip) # if not exist/up-to-date then make()
- > > def exist() # does the target exist?
- > > def uptodate() # is each dependency up to date?

Fully recursive

Persistency dbobject

```
Class DBMeta                                  # python<->db
    def __new__   # makes any derived Class persistent
    def __call__     # instantiate persistent object - attributes

Class DBObject:
    __metaclass__= DBMeta
    object_id = persistent('The object identifier', oidtype)  #unique

# make it
    example = DBObject()
    example.commit()

# get it
    oid = example.object_id
    result = DBObject(object_id = oid)
```

Persistency dataobject

```
from astro.database.DBMain import DBObject, persistent
class DataObject(DBObject):
   filename = persistent('File part of this object',
   str, '')

example =DataObject(pathname='example.txt')
example.store()
example.commit()

g = DataObject.filename.like('example*')
```

Quality control- flags

- Quality flags
  - System quality_flags (method verify)
  - User is_valid (method inspect by user)
    - 0 = bad
    - 1 = OK
    - 2 = Qualified - ready for delivery
- Context - privileges
  - 1 Mydb — user_CalFile
  - 2 Project, eg KIDS
    - Project favorite flag — project CalFile
  - 3 AstroWise — Awe Calfile
  - 4 World
  - 5 VO

Quality control

- Distributed
- Shared over the whole community
- web based
- OmegaCAM calibration plan
- OmegaCAM observing strategies

Example 5LS

#Find ScienceFrames for a ccd named ccd53 and filter

```
Awe> q = (ReducedScienceFrame.chip.name == 'ccd53‘) and
   (ReducedScienceFrame.filter == ‘#841’)
```

From the query result, get the rms of the sky in image

```
Awe> x = [k.imstat.stdev for k in q]
```

get the rms of the used Masterflat

```
Awe> y = [k.flat.imstat.stdev for k in q]
```

Make a plot

```
Awe> pylab.scatter(x,y)
```

the avalanche

- on-the fly re-processing for everything
- 5LS: 5 Lines Script Awe> prompt Python
- Trend analysis Awe > prompt
- All dependent bits are traced “tell_me_everything_tool
- Administration for parallel processing -compute GRID
  - Connect to EGEE - Grid
- Global solutions
  - astrometry/photometry
- Build–in workflow
- Fully user tunable – own provided script
- Context: projects/surveys, instruments, mydb
- Publish directly in EURO-VO

Home | AWE Information System | Instruments | AWE projects

www.astro-wise.org

- What is Astro-WISE?
- Using Astro-WISE
- Publications
- Job openings
- Contact
- Team
- Index
- News Mailinglist
- Issues Mailinglist
- External links
- Search

Astronomical Wide-field Imaging System for Europe

a partnership of

- OmegaCEN-NOVA/Kapteyn Institute, Groningen - NL
- Osservatorio Astronomico di Capodimonte, Napoli - I
- Terapix, IAP, Paris - F
- ESO, Garching bei München - D
- Universitäts-Sternwarte München - D
- OmegaCEN-NOVA - NL

co-ordinated by

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: 363 TB
Number of files stored: 1605413
Database accounts: 104
Total queries[1]: 496136

[1] *sum for all databases since their last restart*

Status of services at Astro-WISE nodes

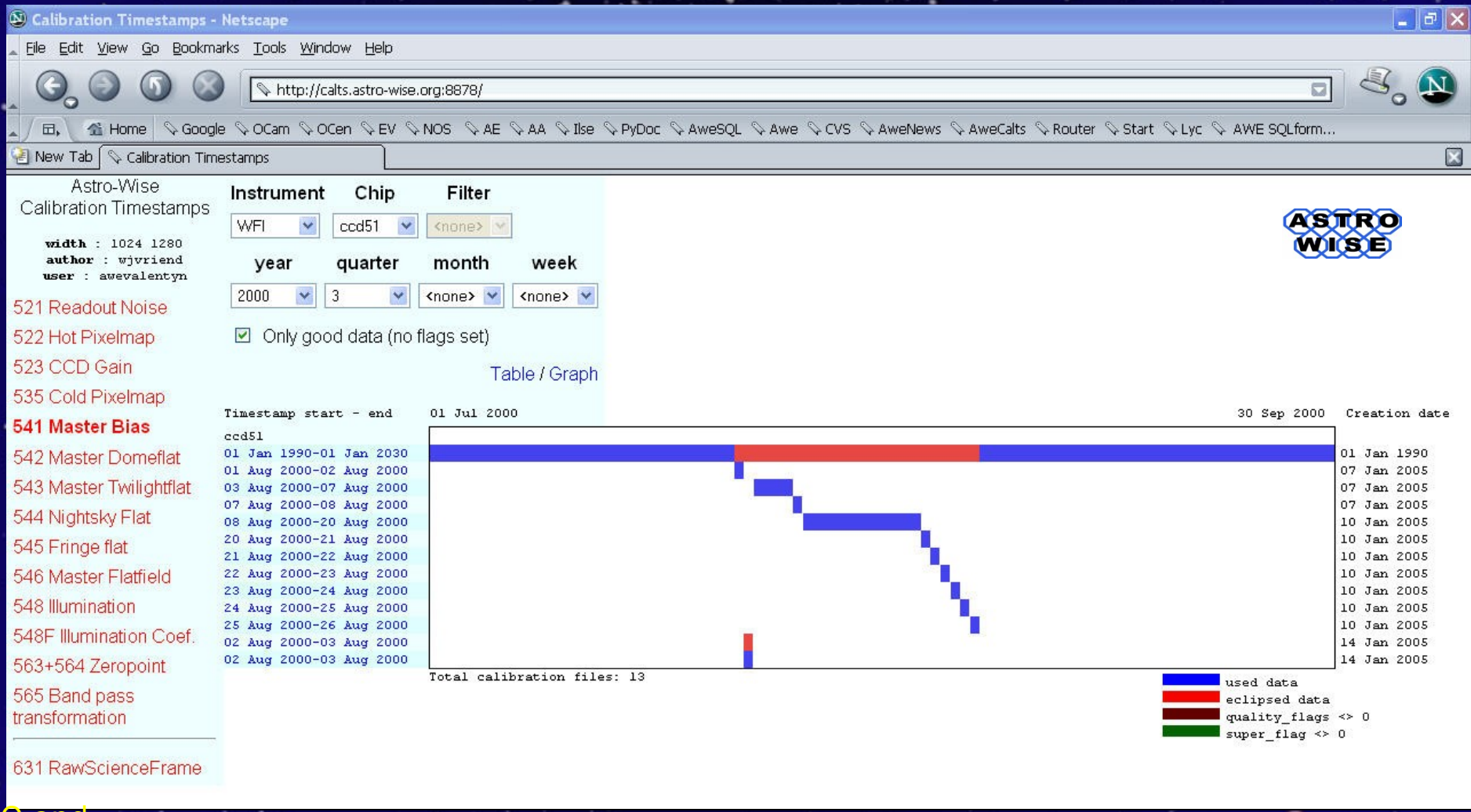
- Bonn
- Groningen
- München
- Napoli

poll details

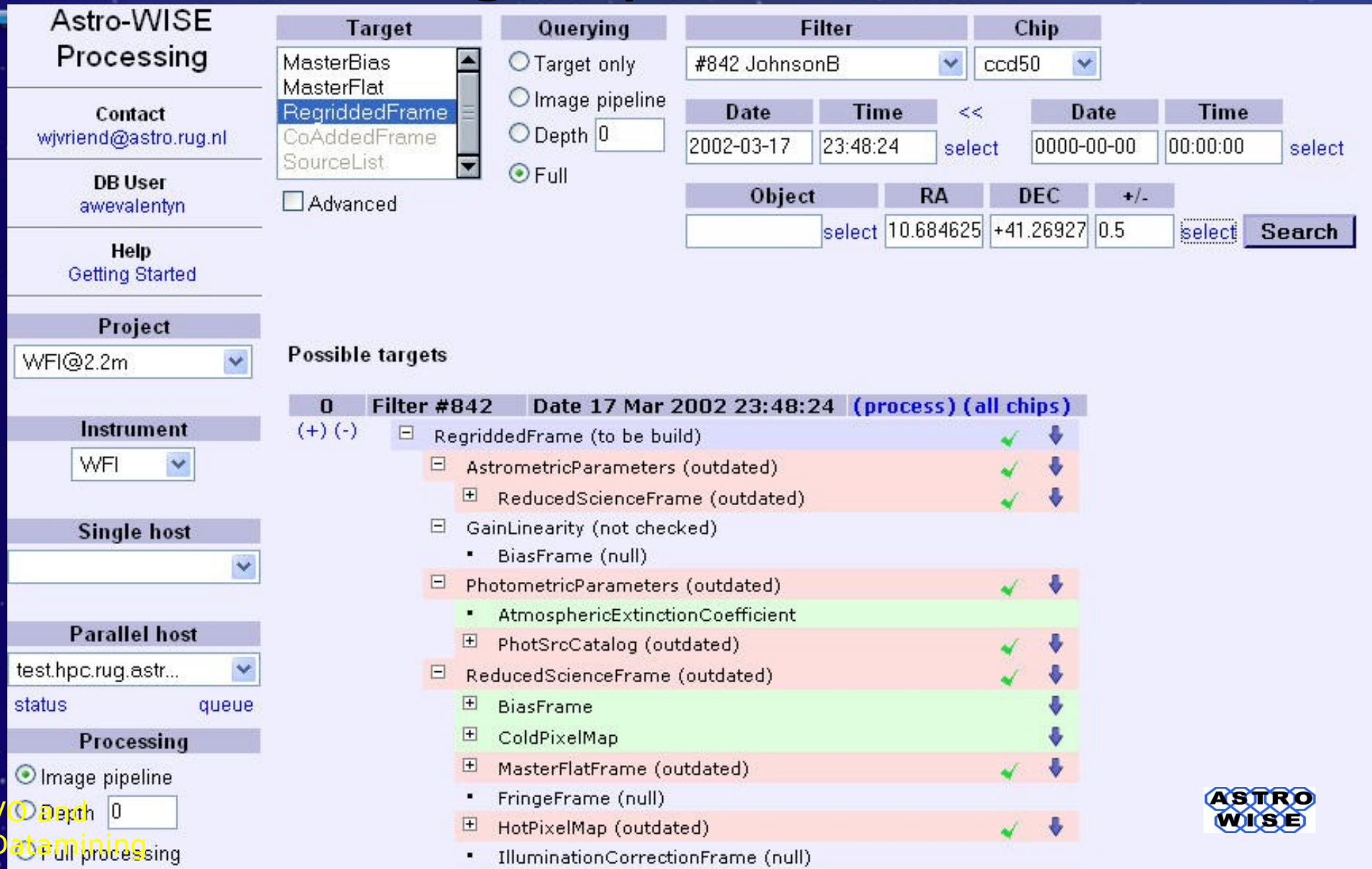
Updated: 14 Jun 2008 18:10:02

VO and Datamining

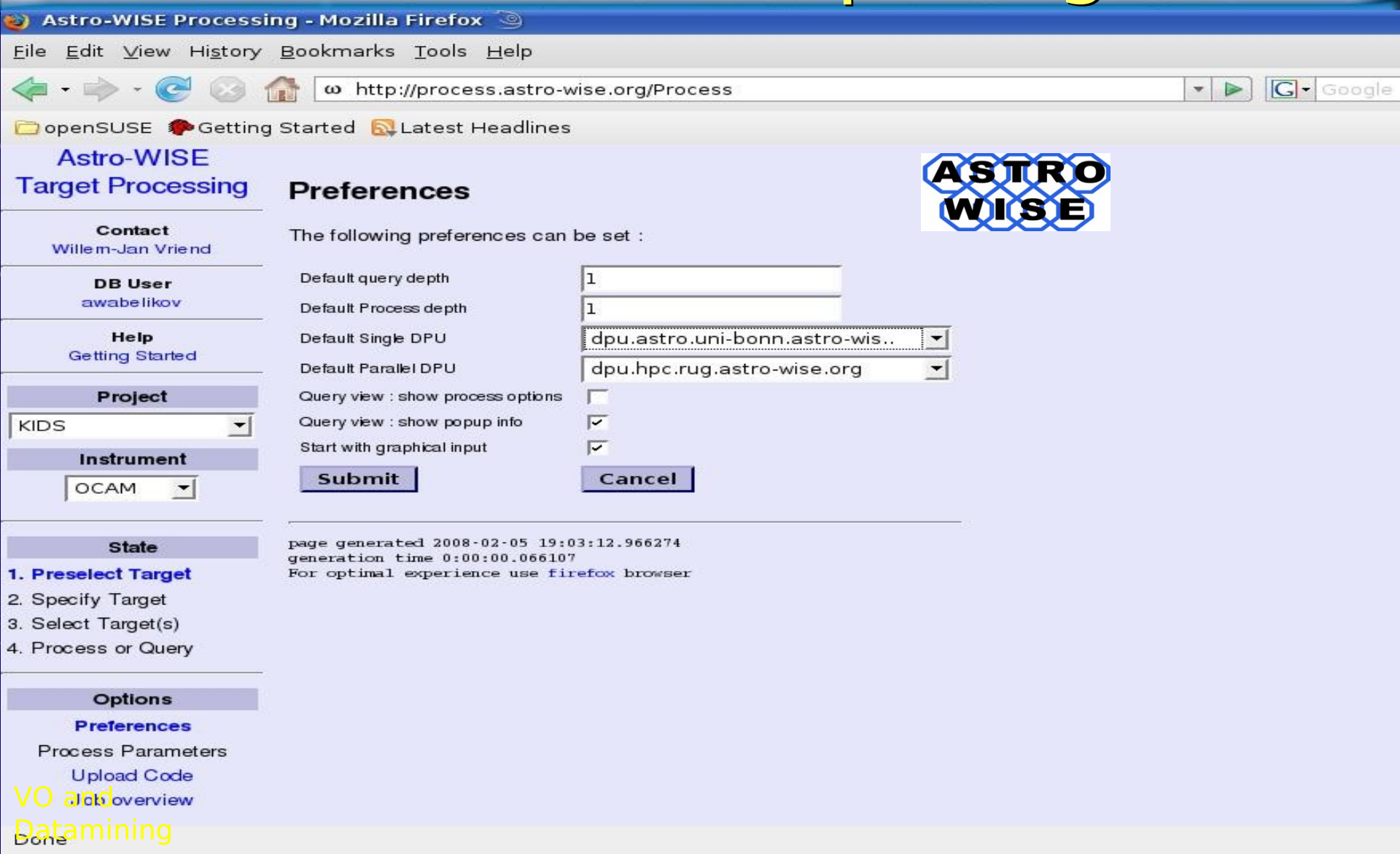
QC - calibration scientist monitoring



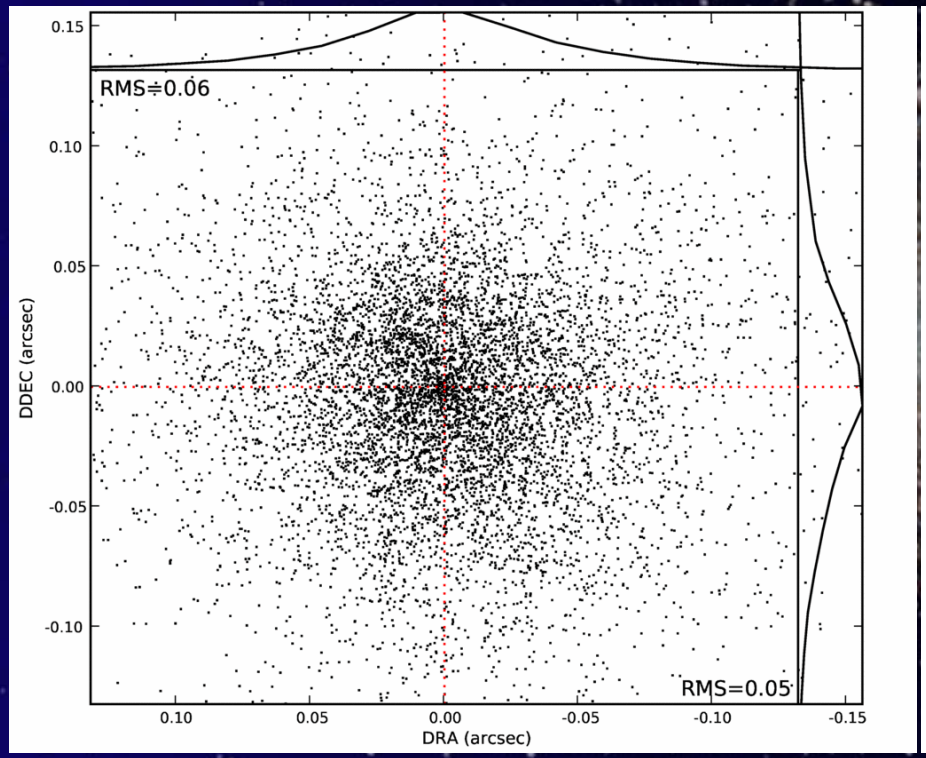
Web services
Target processor

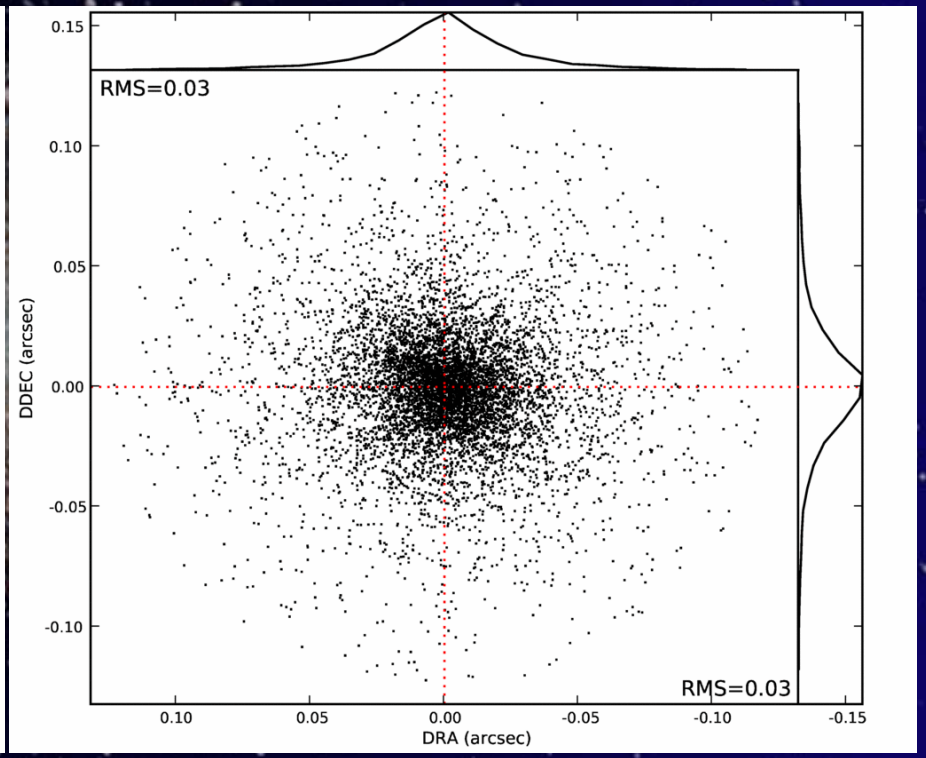


Astro-WISE distributed Computing



Global Astrometry
reproducibility





Local solution

Global solution

(Internal regrid-to-regrid residuals
of one WFI N=4 dither)

File Edit View History Bookmarks Tools Help

http://quality.astro-wise.org/QualityWISE?object_str=RegriddedFrame&object_id=4A5A1A6C90048502E0407D

Google

Quality of RegriddedFrame:

Sci-EVALENTYN-WFI------#842-ccd50-Regr---Sci-54566.3131050-f56144d965b5e765b40bdec3d685fe595215d52b.fits

AstroWISE OBView CalTS Process

no previous comments

DBname: awgvertices project: WFI@2.2m

is_valid = 1: valid | Enter comment her | Submit

Processing Details

creation_date	2008-04-10 07:31:02
is_valid	1
quality_flags	0
Privileges	4

Image Statistics Details

mean	+9.327e+01
median	+8.887e+01
stdev	+2.945e+03
min	-6.036e+06
max	+3.727e+06

Local Astrometry Details

creation_date	2008-04-10 07:30:42
is_valid	1
quality_flags	0
RMS	0.252
SEEING	0.856
NREF	317
SIG_DRA	0.209
SIG_DDEC	0.178
MEAN_DRA	-0.001
MEAN_DDEC	0.002

Photometry Details

creation_date	2008-03-29 20:15:06
is_valid	1
quality_flags	0
zeropoint	24.759
zp_error	0.000
zp_origin	derived
num_sources	173

Observational Details

DATE_OBS	2002-03-18 03:35:21	OBSERVER	UNKNOWN
MJD_OBS	52351.1495509	EXPTIME	299.9176
OBJECT	Sne-W	AIRMSTRT	1.246
R.A.	13:25:30.0000	AIRMEND	1.246
Dec.	-31:35:44.6808	Filter	#842
		mag_id	JohnsonB

Chip ccd50 of Instrument WFI

RegriddedFrame
2439 X 4873 pixel
8.13 X 16.24 arcmin

WeightFrame
2439 X 4873 pixel
8.13 X 16.24 arcmin

AstrometricParameters

Done

Infrared' Eyes i... | Featured Image: Coronet -- A Star-Formatio... | What's Happening: Cosmic Neon Lights the ... | NASA Space Telescopes Find 'Lego-Block' Gal... | Water Vapor Seen 'Raining Do

VO and Datamining

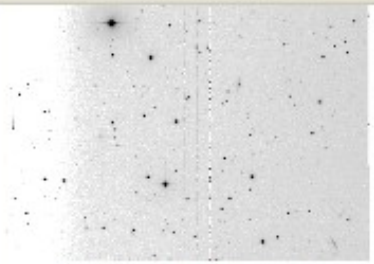
Edit View History Bookmarks Tools Help

http://quality.astro-wise.org/QualityWISE?object_str=RegriddedFrame&object_id=4A5A1A6C90048502E0407D

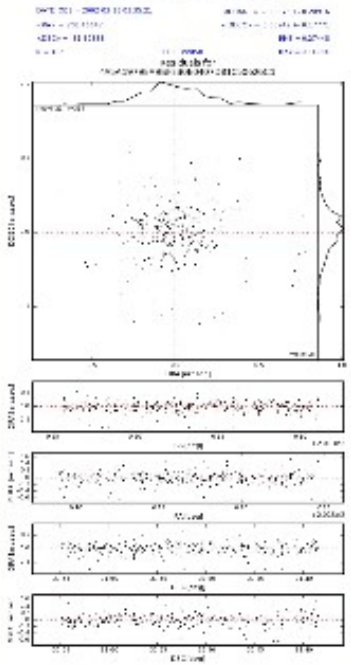
Google

Photometry Details

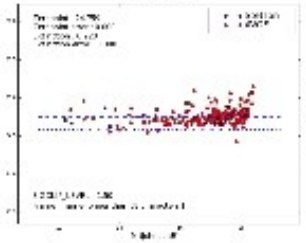
creation_date	2008-03-29 20:15:06
is_valid	1
quality_flags	0
zeropoint	24.759
zp_error	0.000
zp_origin	derived
num_sources	173
extinction	0.220
ext_error	0.000



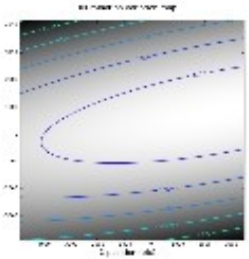
AstrometricParameters



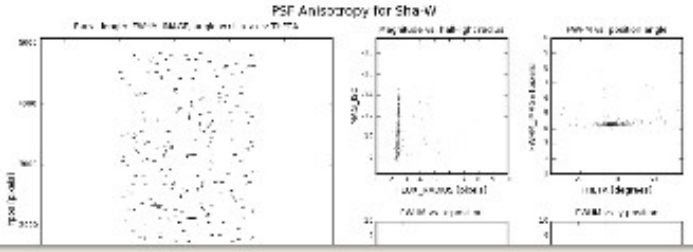
PhotometricParameters



IlluminationCorrection



PSF Anisotropy



ASTRO WISE

VO and Datamining

Supported data sources

Here you can find quick links to different subsets of all public data in our database.

WFI@2.2m data

raw science | reduced science | coadded science | sourcelists | world

Find data for the WFI instrument.

MEGACAM@CFHT data

raw science | reduced science | coadded science | sourcelists | world

Find data for the MegaCAM instrument.

WFC@INT data

raw science | reduced science | coadded science | sourcelists | world

Wide-Field Camera on the Isaac Newton Telescope at La Palma.

SUP@Subaru data

raw science | reduced science | world

Suprime-Cam data from the Subaru telescope on Mauna Kea, Hawaii.

HST ACS data

reduced science | sourcelists | world

Data from the Advanced Camera for Surveys instrument aboard the Hubble Space Telescope.
Only Drizzled images (in the form of ReducedScienceFrames) are currently supported. See this note on ACS data for more information.

SDSS DR7 data

SDSS-Photoz-DR7 | SDSS-SpecObjAll-DR7 | SDSS-PhotoObjAll-DR7 | world

Browse the SDSS DR7 catalog locally as SourceLists.

2MASS PSC data

2MASS PSC SourceList | world

The 2MASS Point Source Catalog is available as a SourceList in our database.

UKIDSS DR3

UKIDSS DR3 SourceList | world

The WFCAM Science Archive Large Area Survey sources are available as a SourceList in our database.

USNO-B1.0

USNO-B1.0 SourceList | world

US Naval Observatory B1.0 catalog, accessable as a SourceList.

+ WFCAM, LBC, ISAAC, **LOFAR, VISTA**

VO and Datamining

Identifying z>5.8 QSO candidates in CFHTLS+VIKING

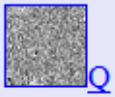
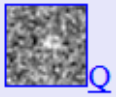
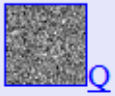
Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help

http://www.astro.rug.nl/~bout/cutout/cutouts142-sel.html

Most Visited | Google Calendar | The Dark Energy Surve... | start [OmegaCEN] | ARTICLES | ABSTRACTS | de.arXiv.org e-Print ar... | Groningen Graduate S...

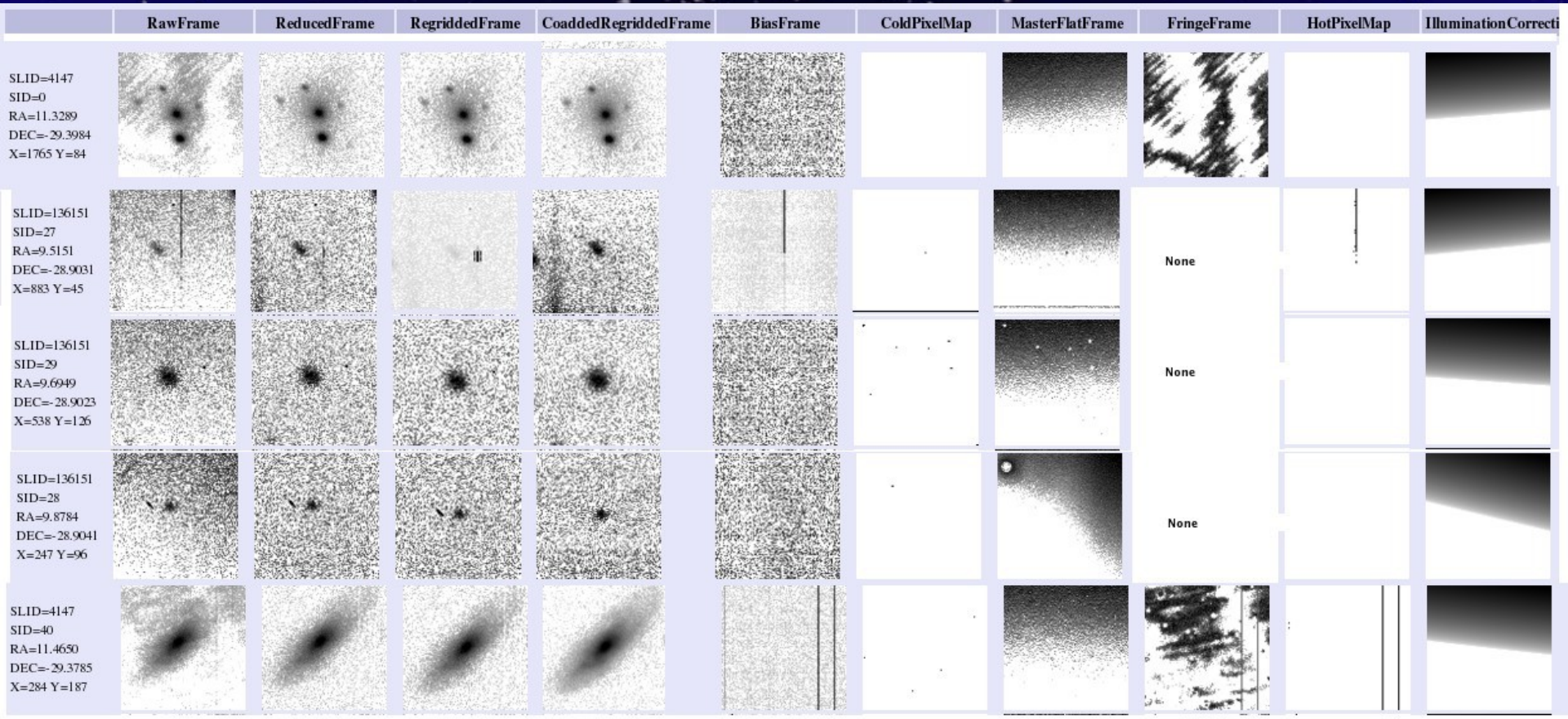
Sources Cutouts

Sources Cutouts

	u	g	r	i	z	Z	Y	J	H	K
Source: (1598441, 12197) RA: 35.51481, DEC: -6.53301	 Q	Q	Q	Q D	Q D	 Q D	Q D	Q D	Q D	Q D
Source: (1598451, 8418) RA: 36.12444, DEC: -6.57637	Q	Q	Q	Q D	Q D	Q D	Q D	Q D	Q D	Q D
Source: (1598841, 71665) RA: 36.27163, DEC: -6.34563	 Q	Q	Q	Q D	Q D	Q D	Q D	Q D	Q D	Q D
Source: (1598951, 66841)										

Klaar

Extreme data lineage

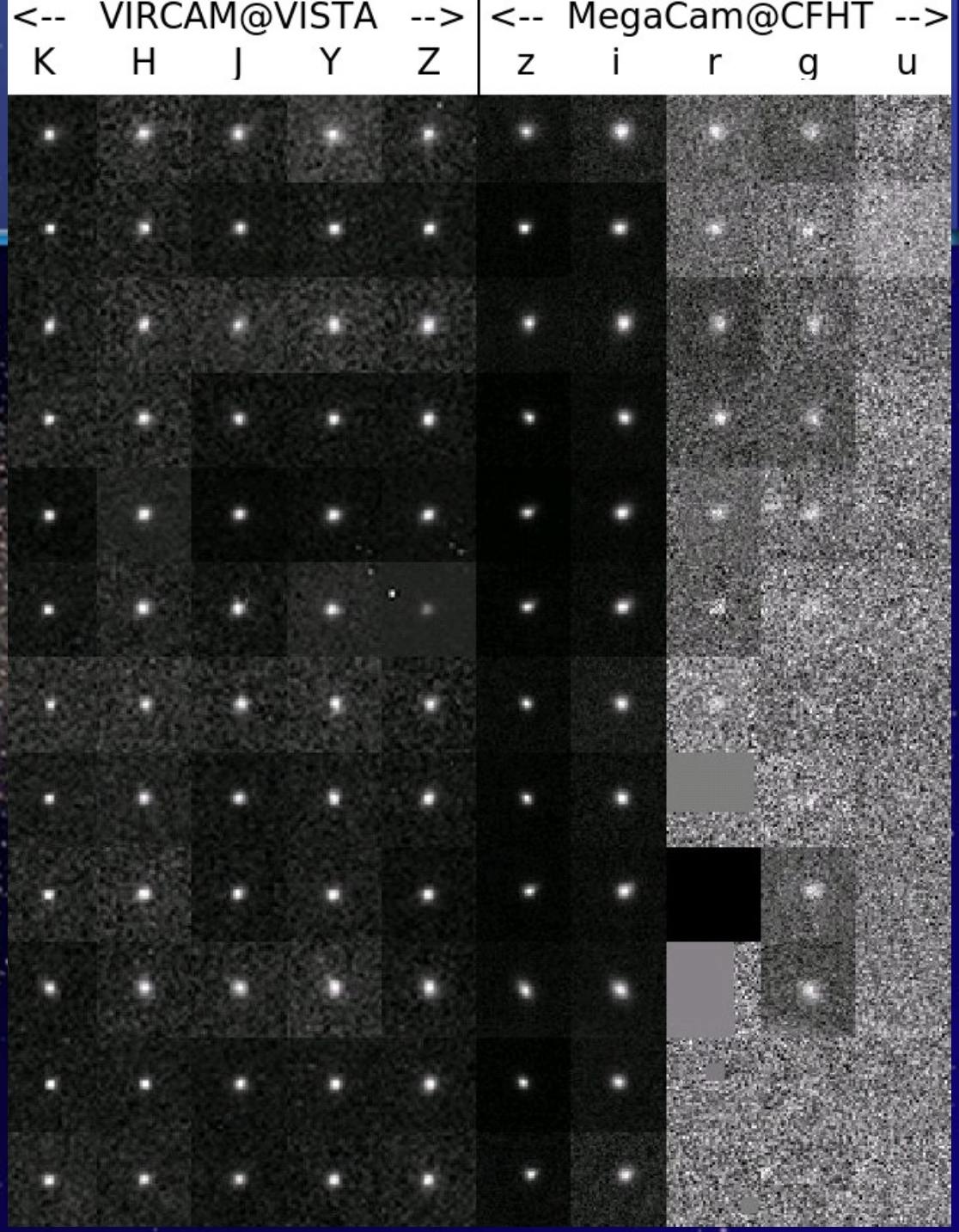


Comp. science journals

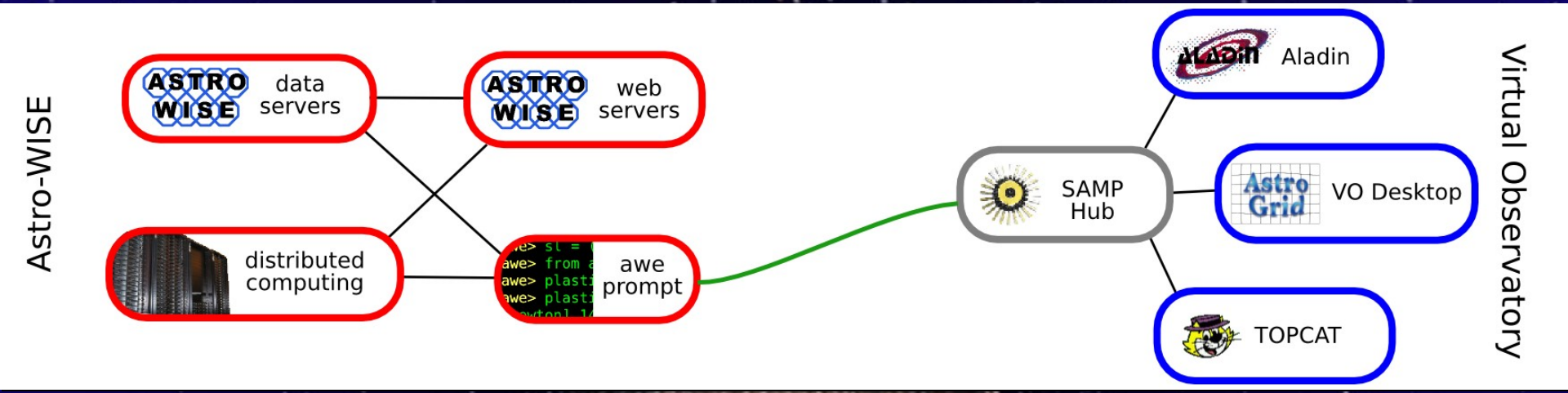
VO and Datamining

Drop-outs

PhD Bout

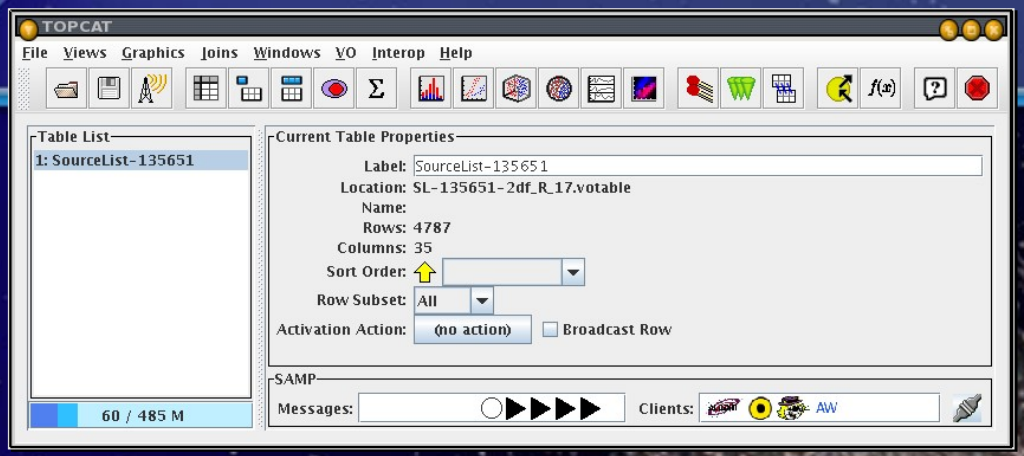


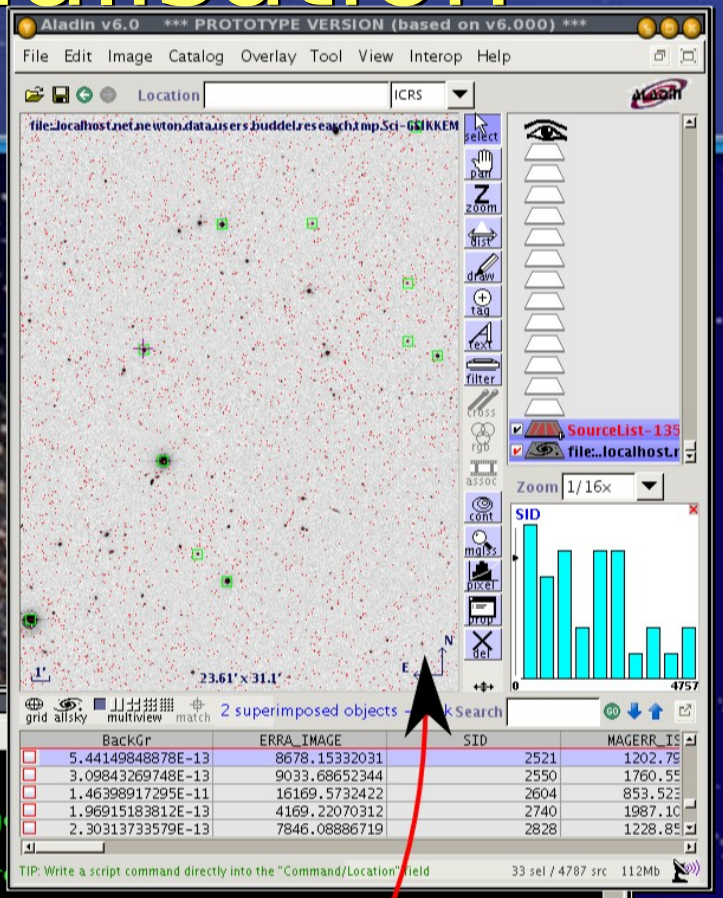
Query driven visualization

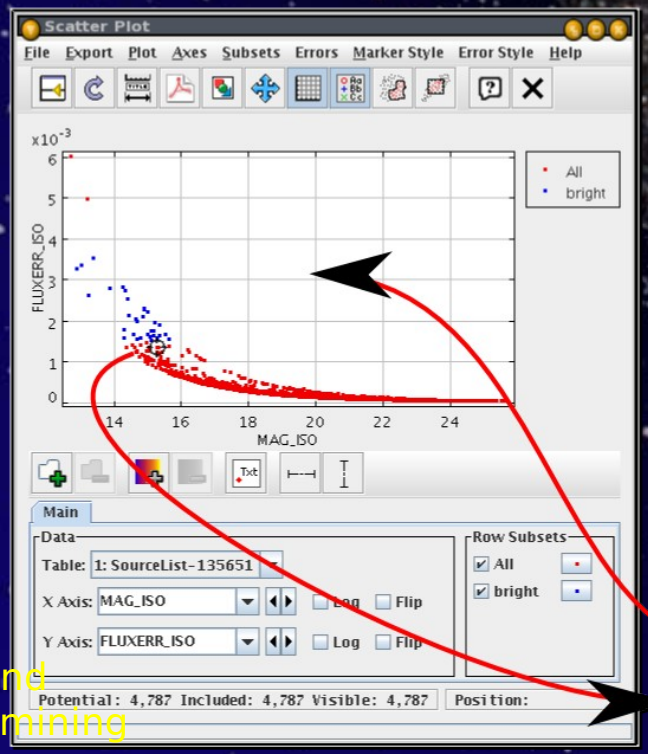


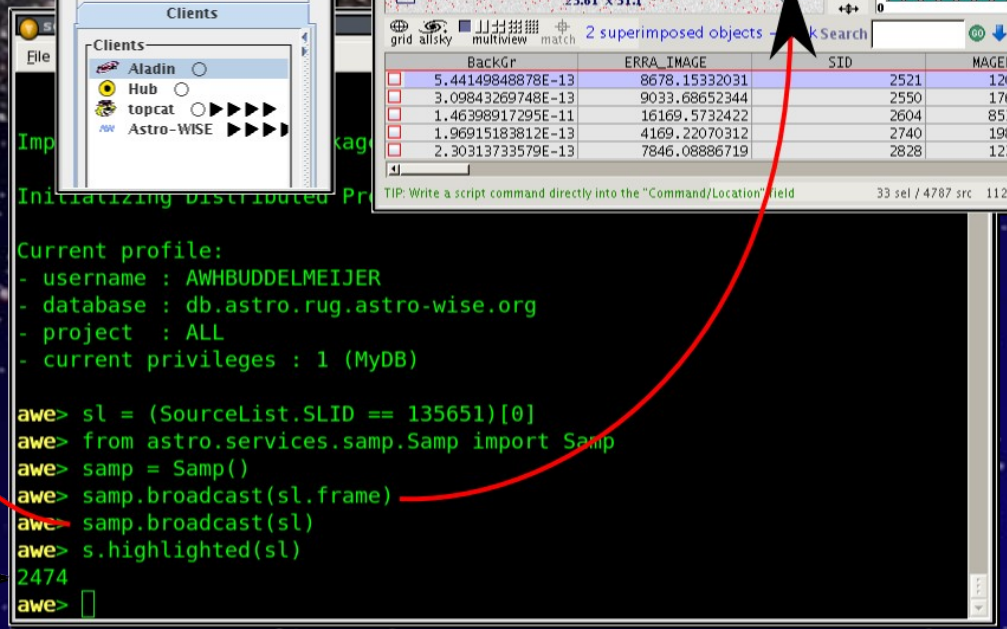
Ph D Buddelmeijer

Query driven visualisation









VO and Datamining

Central role db

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

Astro-WISE paradigm

“Classical” paradigm Forward chaining	**Target processing - Awe Backward chaining**
waterfall model TIER architecture	User hunts upstream
driven by input raw data	Driven by query of user
Process in pipeline workflow	Process in bits and pieces on the fly Backward chaining
Operators push data	User pulls data
Results in releases	Provide information system
Static archives – publish	Dynamic archives –publish Internet
Raw data - obsolete	Raw data is sacred