

# 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](http://www.lofar.org)
- VST-OMEGACAM 2011 [www.astro.rug.nl/~omegacam](http://www.astro.rug.nl/~omegacam)
- GAIA 2012
- ALMA 2013 [www.eso.org/sci/facilities/alma](http://www.eso.org/sci/facilities/alma)
- EUCLID 2018
- LSST 30Tb/night [www.lsst.org](http://www.lsst.org)
- SKA 2022 [www.skatelescope.org](http://www.skatelescope.org)

# Basics- Surveys

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

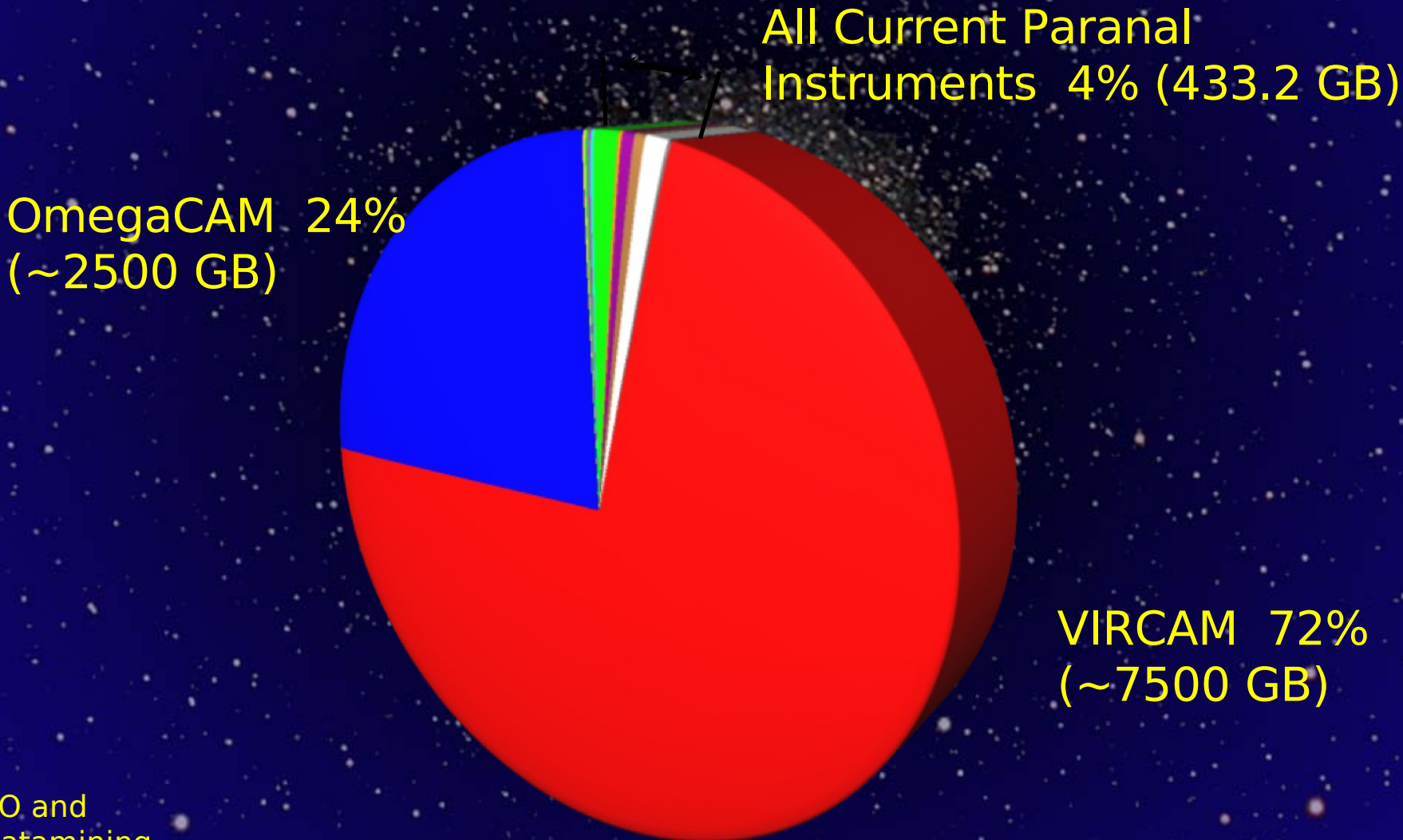
# Basics - time

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  $\leftrightarrow$  data servers  
all Metadata  $\leftrightarrow$  database

compute clusters / GRIDs all I/O to db

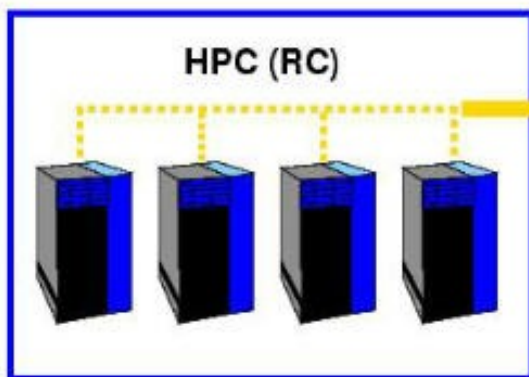
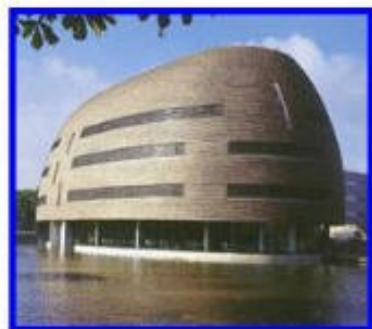
- all components scalable
- all components EU distributed



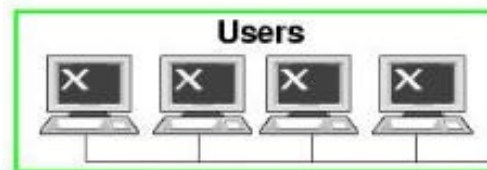
# N params N data back to basics

- Joins – links
- ++ Inheritance – dependencies
- Everything in cs is addresses  
memory, ASCII, namespaces,  
registry
- Optimize , organize, index
- management

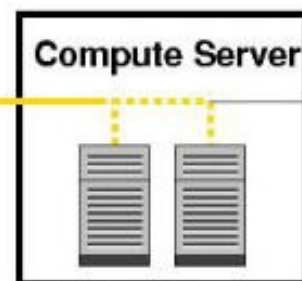
# VST - Virtual Survey Telescope



**HPC (RC)**  
Parallel Pipeline (Python)  
Oracle Client  
FileServer Client (Python)



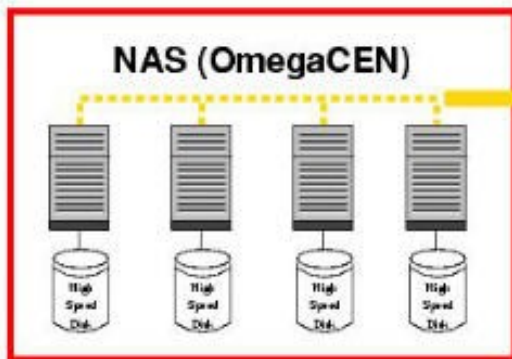
**Users**  
Gateway to Astro-Wise Compute Server



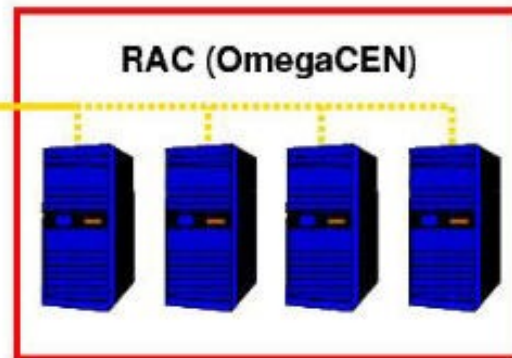
**Compute Server**  
AWE Monitor  
Pipeline (Python)  
Oracle Client  
FileServer Client (Python)

*Leiden  
München  
Napoli  
Paris*

WAN



**NAS (OmegaCEN)**  
FileServer Server (Python)



**RAC (OmegaCEN)**  
Oracle Server



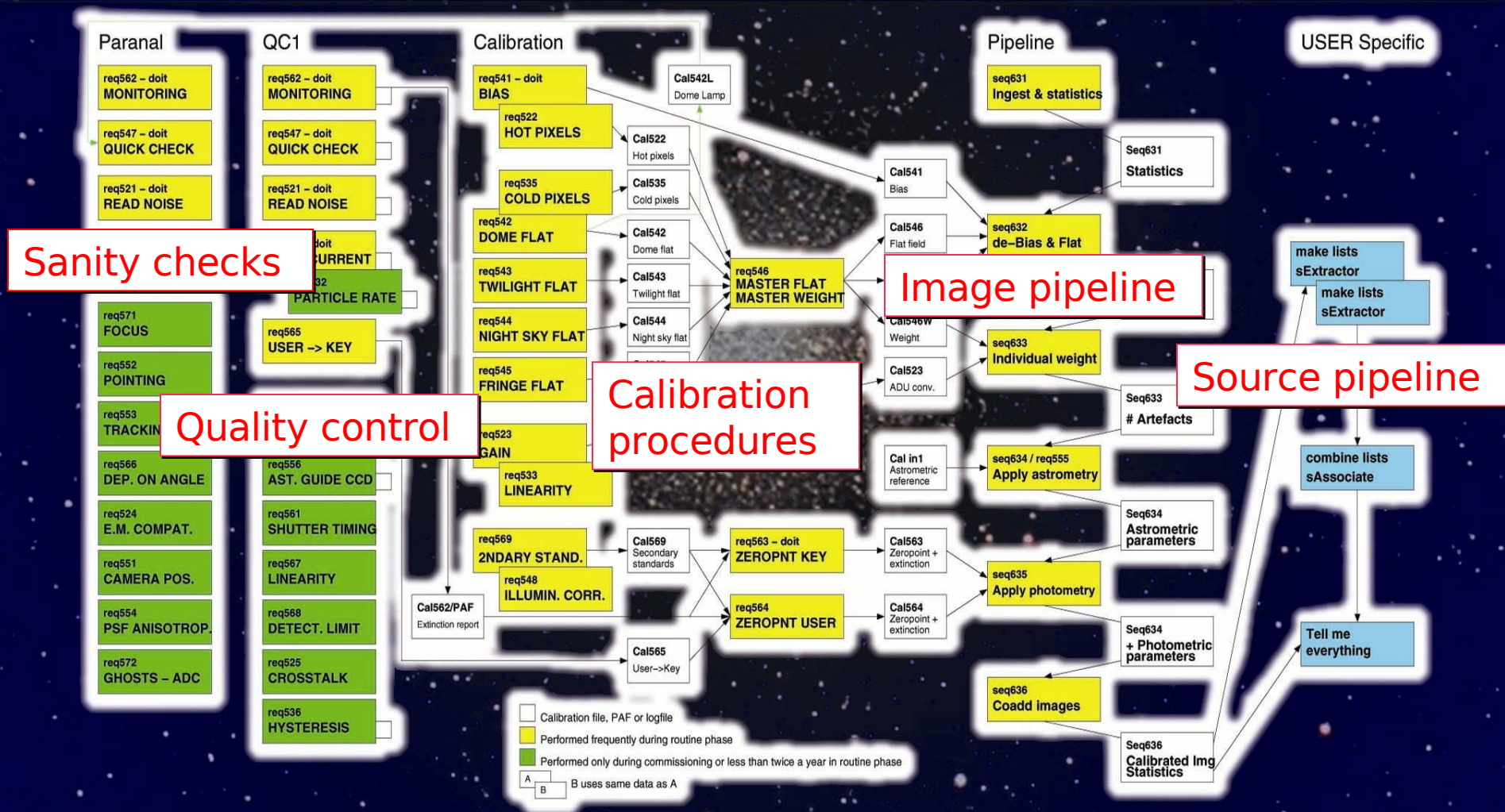
# Centers – satellites



# basics- pipelines

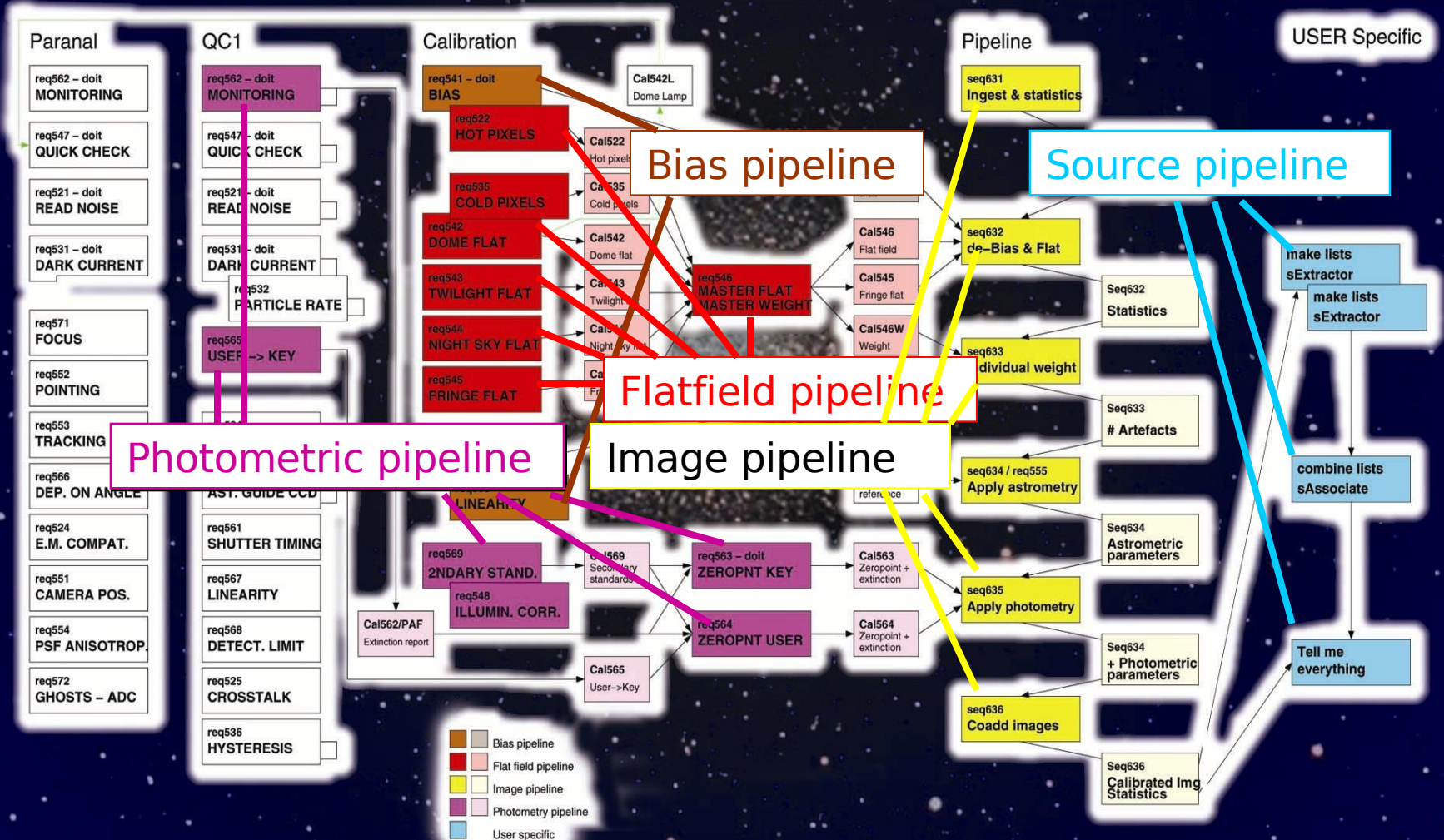
- Workflow
- What triggers a pipeline?
  - Data items
  - Operators
  - users

# basics- Data Model / flow

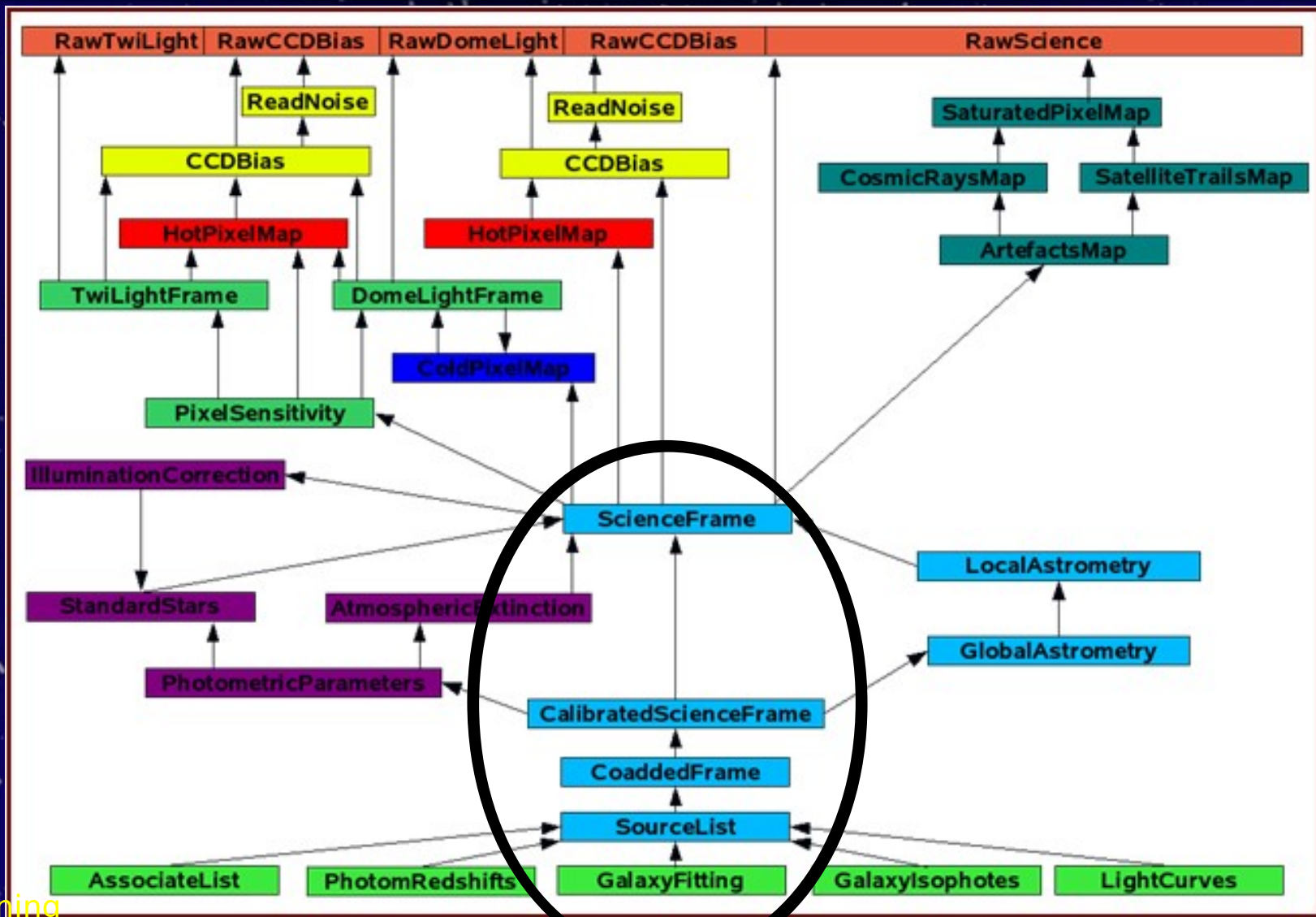


# Basics

## 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 CalFile
  - Project favorite flag Awe Calfile
- 3 AstroWise
- 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



www.astro-wise.org



## 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<sup>1</sup>: 496136

<sup>1</sup>sum for all databases since their last restart

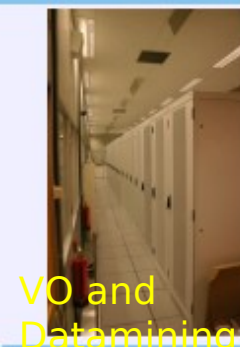
### Status of services at Astro-WISE nodes

[Bonn](#)   
[Groningen](#)   
[München](#)   
[Napoli](#)

[poll details](#)

Updated: 14 Jun 2008 18:10:02

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



VO and Datamining

# QC - calibration scientist monitoring

Calibration Timestamps - Netscape

File Edit View Go Bookmarks Tools Window Help

http://calts.astro-wise.org:8878/

Home Google OCam OCen EV NOS AE AA Ise PyDoc AweSQL Awe CVS AweNews AweCalts Router Start Lyc AWE SQLform...

New Tab Calibration Timestamps

Astro-Wise Calibration Timestamps

width : 1024 1280  
author : wjvriend  
user : awevalentyn

521 Readout Noise  
522 Hot Pixelmap  
523 CCD Gain  
535 Cold Pixelmap  
**541 Master Bias**  
542 Master Domeflat  
543 Master Twilightflat  
544 Nightsky Flat  
545 Fringe flat  
546 Master Flatfield  
548 Illumination  
548F Illumination Coef.  
563+564 Zeropoint  
565 Band pass transformation  
631 RawScienceFrame

Instrument: WFI Chip: ccd51 Filter: <none>

year: 2000 quarter: 3 month: <none> week: <none>

Only good data (no flags set)

Table / Graph

Timestamp start - end 01 Jul 2000 30 Sep 2000 Creation date

ccd51

Timestamp start - end	Creation date
01 Jan 1990-01 Jan 2030	01 Jan 1990
01 Aug 2000-02 Aug 2000	07 Jan 2005
03 Aug 2000-07 Aug 2000	07 Jan 2005
07 Aug 2000-08 Aug 2000	07 Jan 2005
08 Aug 2000-20 Aug 2000	10 Jan 2005
20 Aug 2000-21 Aug 2000	10 Jan 2005
21 Aug 2000-22 Aug 2000	10 Jan 2005
22 Aug 2000-23 Aug 2000	10 Jan 2005
23 Aug 2000-24 Aug 2000	10 Jan 2005
24 Aug 2000-25 Aug 2000	10 Jan 2005
25 Aug 2000-26 Aug 2000	10 Jan 2005
02 Aug 2000-03 Aug 2000	14 Jan 2005
02 Aug 2000-03 Aug 2000	14 Jan 2005

Total calibration files: 13

Legend:

- used data
- eclipsed data
- quality\_flags <> 0
- super\_flag <> 0



# Web services Target processor

### Astro-WISE Processing

---

**Contact**  
wjvriend@astro.rug.nl

---

**DB User**  
awevalentyn

---

**Help**  
Getting Started

---

**Project**  
WFI@2.2m

---

**Instrument**  
WFI

---

**Single host**

---

**Parallel host**  
test.hpc.rug.astr...

status      queue

---

**Processing**

Image pipeline  
Depth 0

Full processing

Target	Querying	Filter	Chip												
MasterBias MasterFlat <b>RegriddedFrame</b> CoAddedFrame SourceList	<input type="radio"/> Target only <input type="radio"/> Image pipeline <input type="radio"/> Depth 0 <input checked="" type="radio"/> Full	#842 JohnsonB	ccd50												
<input type="checkbox"/> Advanced		<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Date</th> <th>Time</th> <th>&lt;&lt;</th> <th>Date</th> <th>Time</th> <th>select</th> </tr> </thead> <tbody> <tr> <td>2002-03-17</td> <td>23:48:24</td> <td></td> <td>0000-00-00</td> <td>00:00:00</td> <td>select</td> </tr> </tbody> </table>	Date	Time	<<	Date	Time	select	2002-03-17	23:48:24		0000-00-00	00:00:00	select	
Date	Time	<<	Date	Time	select										
2002-03-17	23:48:24		0000-00-00	00:00:00	select										
		<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Object</th> <th>RA</th> <th>DEC</th> <th>+/-</th> <th>select</th> </tr> </thead> <tbody> <tr> <td></td> <td>10.684625</td> <td>+41.26927</td> <td>0.5</td> <td>select</td> </tr> </tbody> </table>	Object	RA	DEC	+/-	select		10.684625	+41.26927	0.5	select	<b>Search</b>		
Object	RA	DEC	+/-	select											
	10.684625	+41.26927	0.5	select											

**Possible targets**

0	Filter #842	Date 17 Mar 2002 23:48:24	(process) (all chips)
(+) (-)	<input type="checkbox"/> RegriddedFrame (to be build)		✓ ↓
	<input type="checkbox"/> AstrometricParameters (outdated)		✓ ↓
	<input checked="" type="checkbox"/> ReducedScienceFrame (outdated)		✓ ↓
	<input type="checkbox"/> GainLinearity (not checked)		
	▪ BiasFrame (null)		
	<input type="checkbox"/> PhotometricParameters (outdated)		✓ ↓
	▪ AtmosphericExtinctionCoefficient		
	<input checked="" type="checkbox"/> PhotSrcCatalog (outdated)		✓ ↓
	<input type="checkbox"/> ReducedScienceFrame (outdated)		✓ ↓
	<input checked="" type="checkbox"/> BiasFrame		✓ ↓
	<input checked="" type="checkbox"/> ColdPixelMap		✓ ↓
	<input checked="" type="checkbox"/> MasterFlatFrame (outdated)		✓ ↓
	▪ FringeFrame (null)		
	<input checked="" type="checkbox"/> HotPixelMap (outdated)		✓ ↓
	▪ IlluminationCorrectionFrame (null)		

# Astro-WISE distributed Computing

Astro-WISE Processing - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://process.astro-wise.org/Process

openSUSE Getting Started Latest Headlines

## Astro-WISE Target Processing

**Contact**  
Willem-Jan Vriend

**DB User**  
awabelikov

**Help**  
Getting Started

**Project**  
KIDS

**Instrument**  
OCAM

**State**

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

**Options**

**Preferences**

Process Parameters

Upload Code

Job Overview

Done

## Preferences

The following preferences can be set :

Default query depth: 1

Default Process depth: 1

Default Single DPU: dpu.astro.uni-bonn.astro-wis..

Default Parallel DPU: dpu.hpc.rug.astro-wise.org


Query view : show process options:

Query view : show popup info:

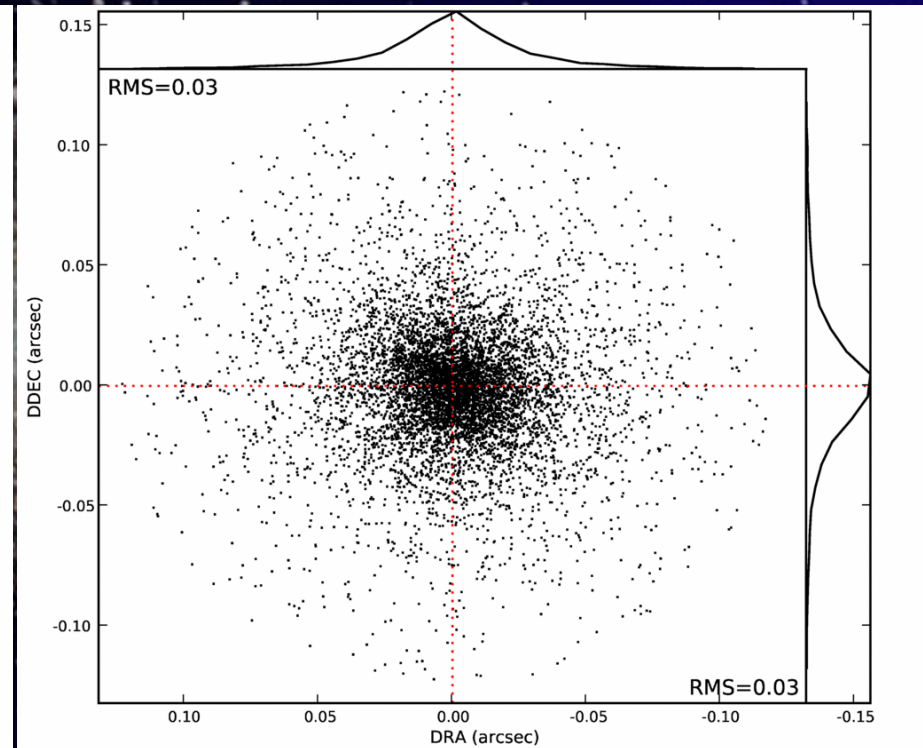
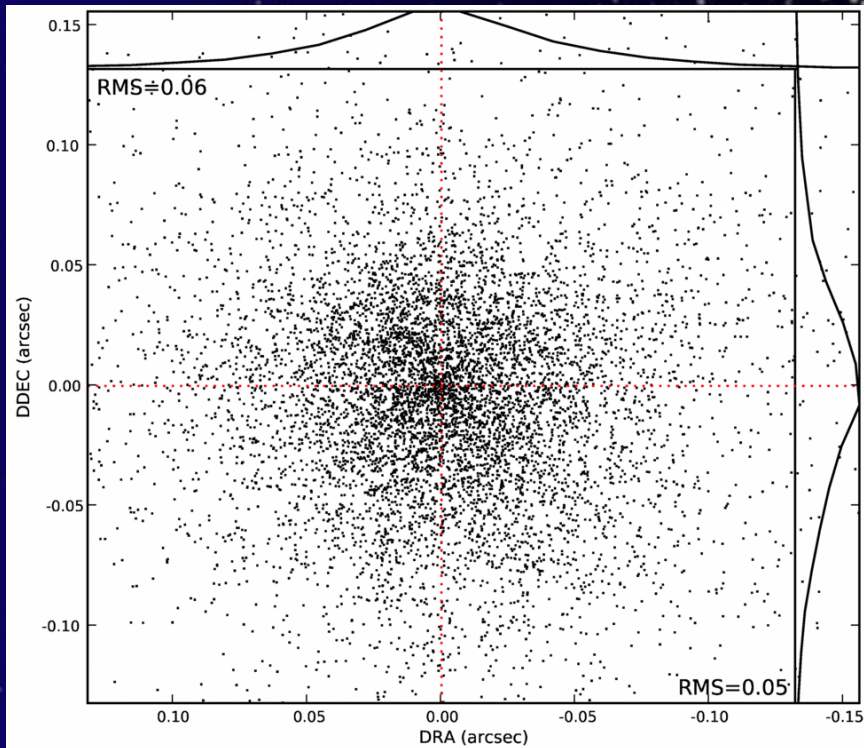
Start with graphical input:

**Submit** **Cancel**

page generated 2008-02-05 19:03:12.966274  
generation time 0:00:00.066107  
For optimal experience use [firefox](#) browser



# Global Astrometry reproducibility



Local solution

Global solution

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

### Quality of REGRIDDEDFRAME:

Sci-EVAL ENTYN-WFI #842-ccd50-Regr--Sci-54566.3131050-f56144d965b5e765b40bdec3d685fe595215d52b.fits

ASTROWISE Overview CALTS Process

no previous comments

OBName: [zgwvdr008](#) project: WFI@2.2m

is\_valid = 1: valid

#### 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

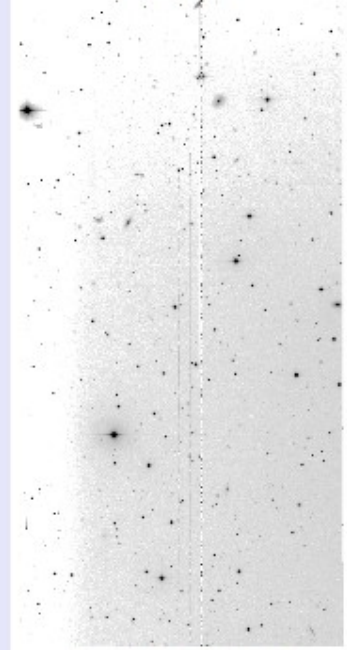
DATE_OBS	2008-03-18 03:35:21	OB_SERVER	UNKNOWN
MJD_OBS	52351.1495509	EXPTIME	299.9176
OBJECT	Bra-W	AIRW_BRT	1.246
R.A.	13:25:30.0000	AIRW_MD	1.246
Dec.	-31:38:44.6908	Filter	#342
		mag_id	JohnsonB

#### Chip ccd60 of Instrument WFI



#### RegriddedFrame

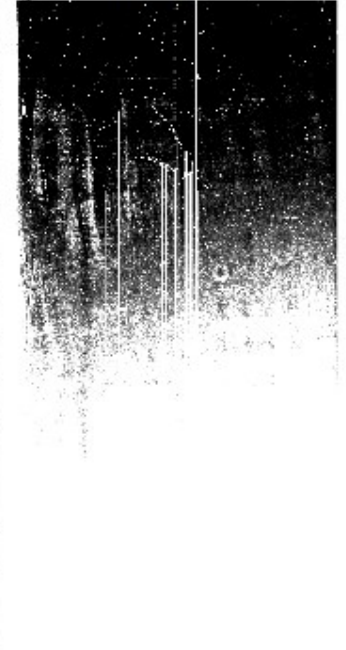
2439 X 4873 pixel  
8.13 X 16.24 arcmin



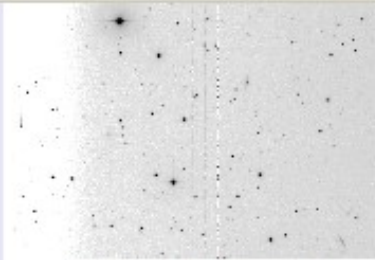
AstrometricParameters

#### WeightFrame

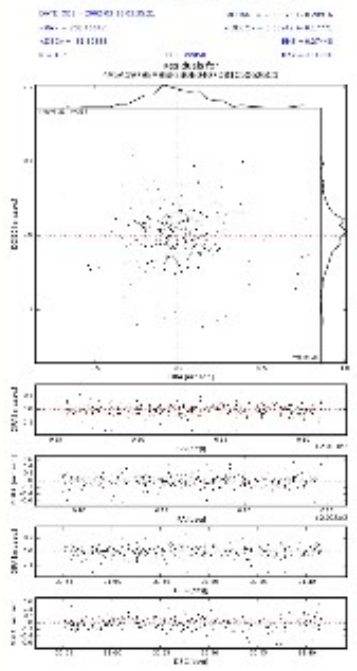
2439 X 4873 pixel  
8.13 X 16.24 arcmin



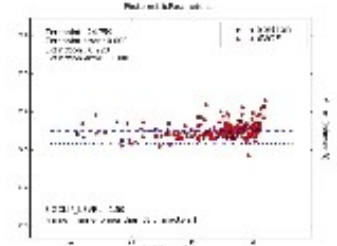
Photometry Details	
creation_date	2008-03-29 20:15:06
is_valid	1
quality_flags	0
zeropoint	24.759
zp_error	0.000
zp_origin	derived
num_sources	173
extinction	0.220
ext_error	0.000



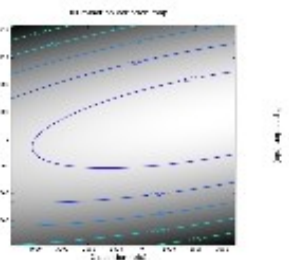
### AstrometricParameters



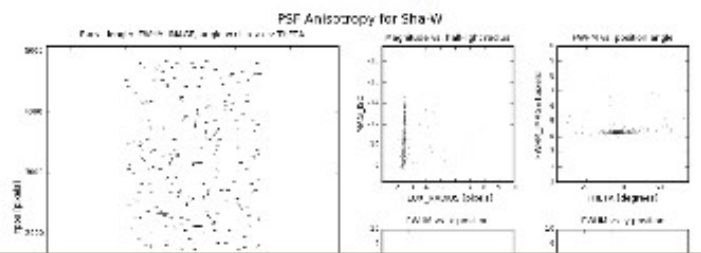
### PhotometricParameters



### IlluminationCorrection



### PSF Anisotropy



# 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

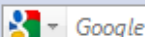
[USNO-B1.0 SourceList](#)


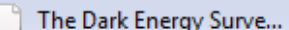
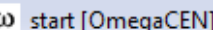
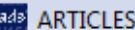
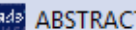
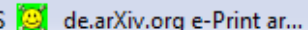
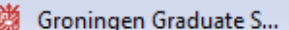
[world](#)

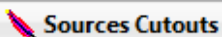
US Naval Observatory B1.0 catalog, accessible as a SourceList.

# Identifying $z > 5.8$ QSO candidates in CFHTLS+VIKING


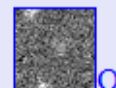
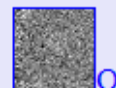
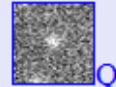
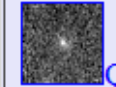





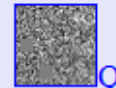
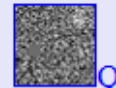
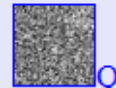
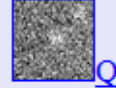
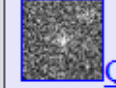
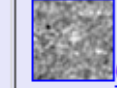




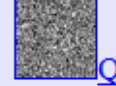
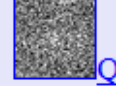

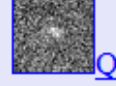







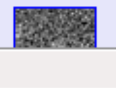







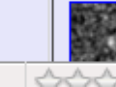
Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help

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

Most Visited  Google Calendar  The Dark Energy Survey  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

	RawFrame	ReducedFrame	RegriddedFrame	CoaddedRegriddedFrame	BiasFrame	ColdPixelMap	MasterFlatFrame	FringeFrame	HotPixelMap	Illumination Correcti
SLID=4147 SID=0 RA=11.3289 DEC=-29.3984 X=1765 Y=84										
SLID=136151 SID=27 RA=9.5151 DEC=-28.9031 X=883 Y=45								None		
SLID=136151 SID=29 RA=9.6949 DEC=-28.9023 X=538 Y=126								None		
SLID=136151 SID=28 RA=9.8784 DEC=-28.9041 X=247 Y=96								None		
SLID=4147 SID=40 RA=11.4650 DEC=-29.3785 X=284 Y=187										

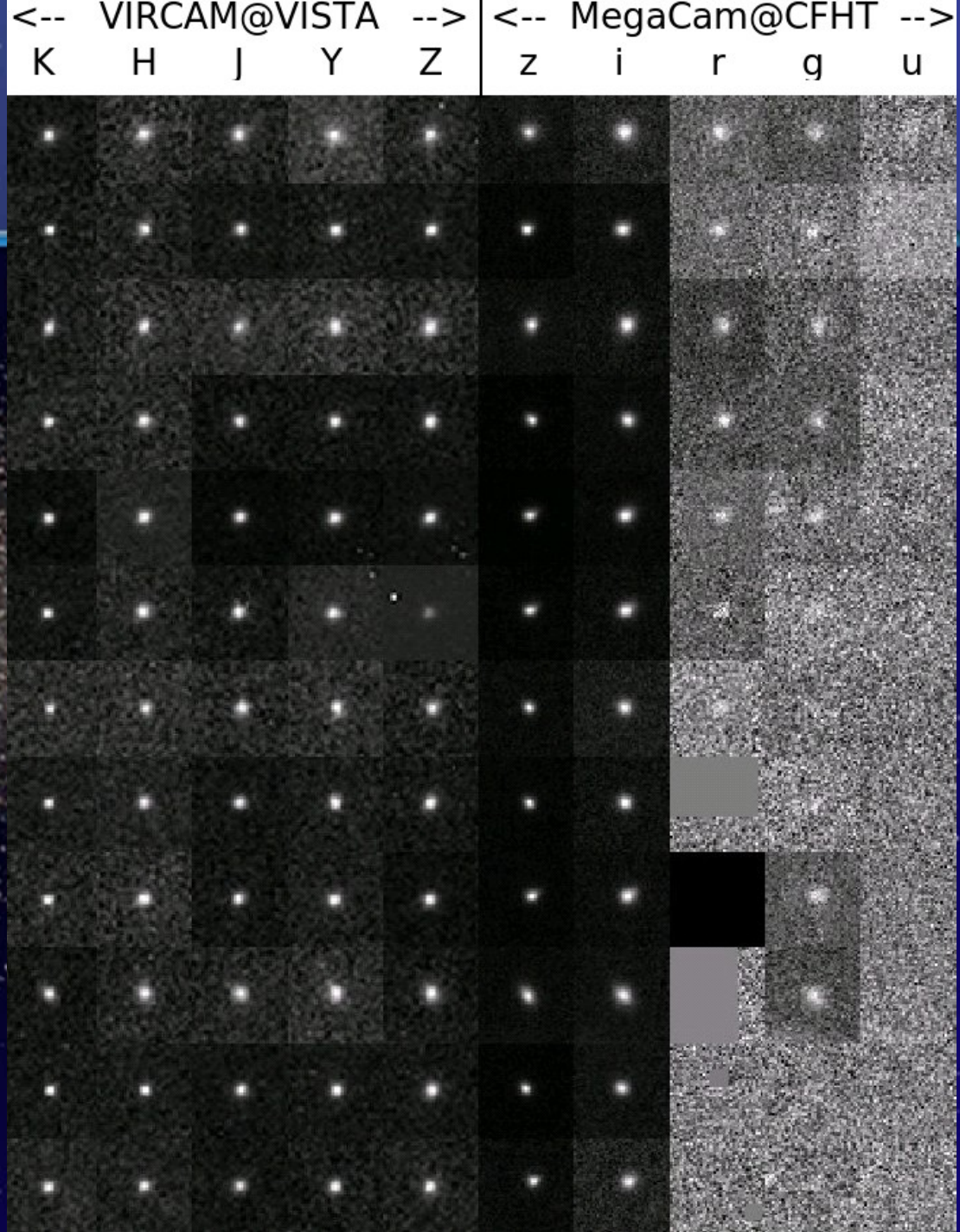
Comp. science journals



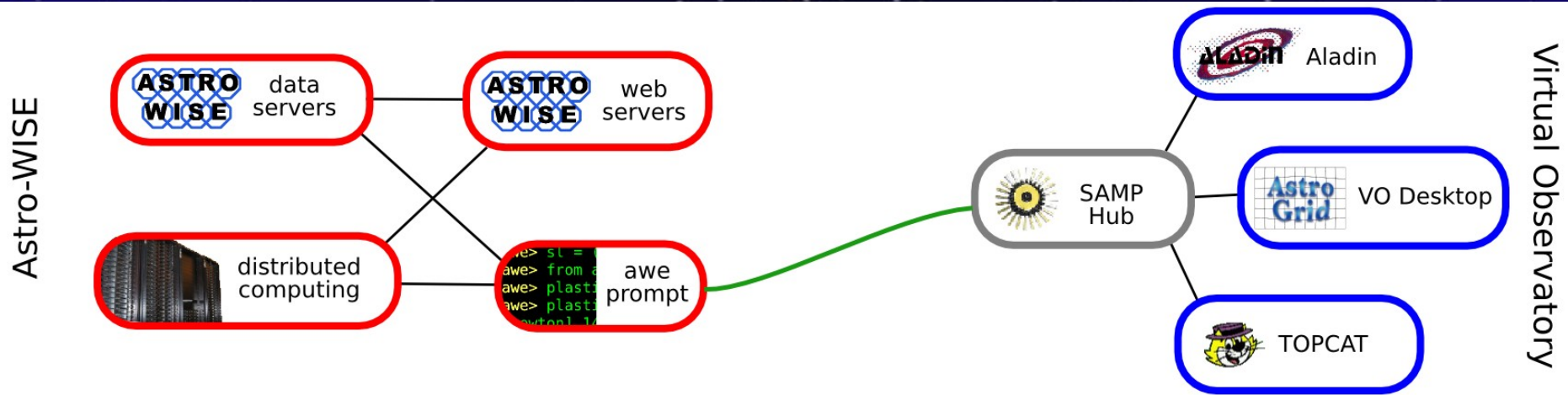
# Drop-outs

PhD Bout

VO and Datamining



# Query driven visualization



Ph D Buddelmeijer

# Query driven visualisation

TOPCAT

File Views Graphics Joins Windows YO Interop Help

Table List

- 1: SourceList-135651

Current Table Properties

Label: SourceList-135651  
Location: SL-135651-2df\_R\_17.votable  
Name:  
Rows: 4787  
Columns: 35  
Sort Order:  $\uparrow$   
Row Subset: All  
Activation Action: (no action)  Broadcast Row

SAMP

Messages:      Clients: AW

60 / 485 M

Aladin v6.0 \*\*\* PROTOTYPE VERSION (based on v6.000) \*\*\*

File Edit Image Catalog Overlay Tool View Interop Help

Location: ICRS

file:localhost.net:awton.data.users.buddelmeijer@mp.sci-6.nl:KEM

2 superimposed objects

BackGr	ERRA_IMAGE	SID	MAGERR_IS
5.44149848878E-13	8678.15332031	2521	1202.79
3.09843269748E-13	9033.68652344	2550	1760.59
1.46398917295E-11	16169.5732422	2604	853.52
1.96915183812E-13	4169.22070312	2740	1987.10
2.30313733579E-13	7846.08886719	2828	1228.89

TIP: Write a script command directly into the "Command/Location" field

33 sel / 4787 src 112Mb

Scatter Plot

File Export Plot Axes Subsets Errors Marker Style Error Style Help

FLUXERR\_ISO (y-axis,  $\times 10^{-3}$ )

MAG\_ISO (x-axis)

Legend: All (red dot), bright (blue dot)

Main

Data

Table: 1: SourceList-135651

X Axis: MAG\_ISO  Log  Flip

Y Axis: FLUXERR\_ISO  Log  Flip

Row Subsets

- All
- bright

Potential: 4,787 Included: 4,787 Visible: 4,787 Position:

SAMP Control

File Connect Help

Sent Messages

Received Messages

Clients

- Aladin
- Hub
- topcat
- Astro-WISE

Current profile:

- username : AWHBUDELMEIJER
- database : db.astro.rug.astro-wise.org
- project : ALL
- current privileges : 1 (MyDB)

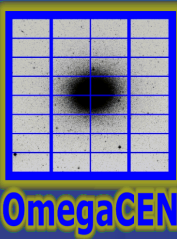
```
awe> sl = (SourceList.SLID == 135651)[0]
awe> from astro.services.samp.Samp import Samp
awe> samp = Samp()
awe> samp.broadcast(sl.frame)
awe> samp.broadcast(sl)
awe> s.highlighted(sl)
2474
awe>
```

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



<b>“Classical” paradigm</b>	<b>Target processing - Awe</b>
<b>Forward chaining</b>	<b>Backward chaining</b>
waterfall model	User hunts upstream
TIER architecture	
driven by input raw data	Driven by query of user
Process in pipeline	Process in bits and pieces on the fly
workflow	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