Interacademic Lecture 8

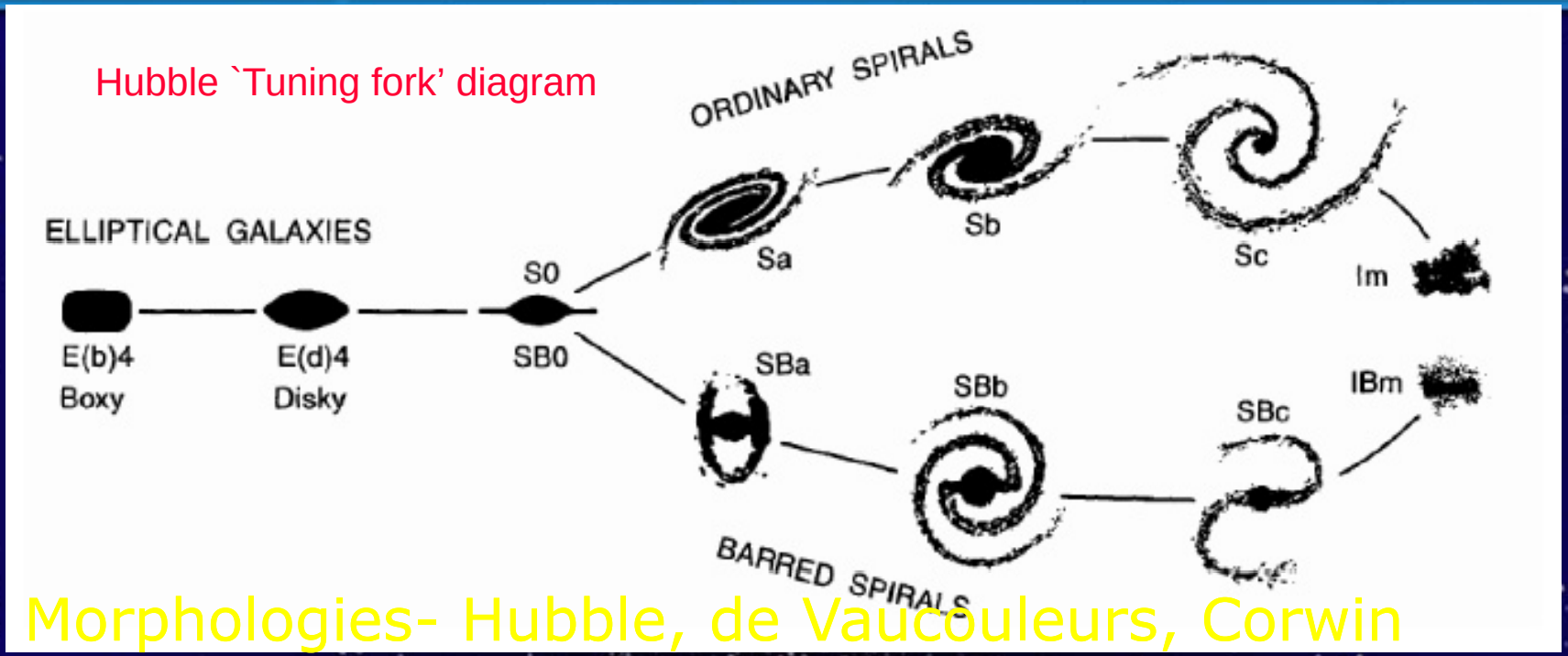
Future directions in Astronomical data management

Kapteyn Astronomical Institute
University of Groningen

Edwin A. Valentijn

12 May 2010

Case: morphologies
parameters vs theory -degeneracy



Morphologies- Hubble, de Vaucouleurs, Corwin

- U, B, V, R , I, Z, K
  - m_tot, m_25, m_26, D_26, D_25, D_90%, r_eff, sb_0, sb_eff
- Structure/radial
  - a/b, pa, B/D, N, exp scale length , delta (B-R)
- colours
  - Central (B-R), (B_R)tot

What characterizes?

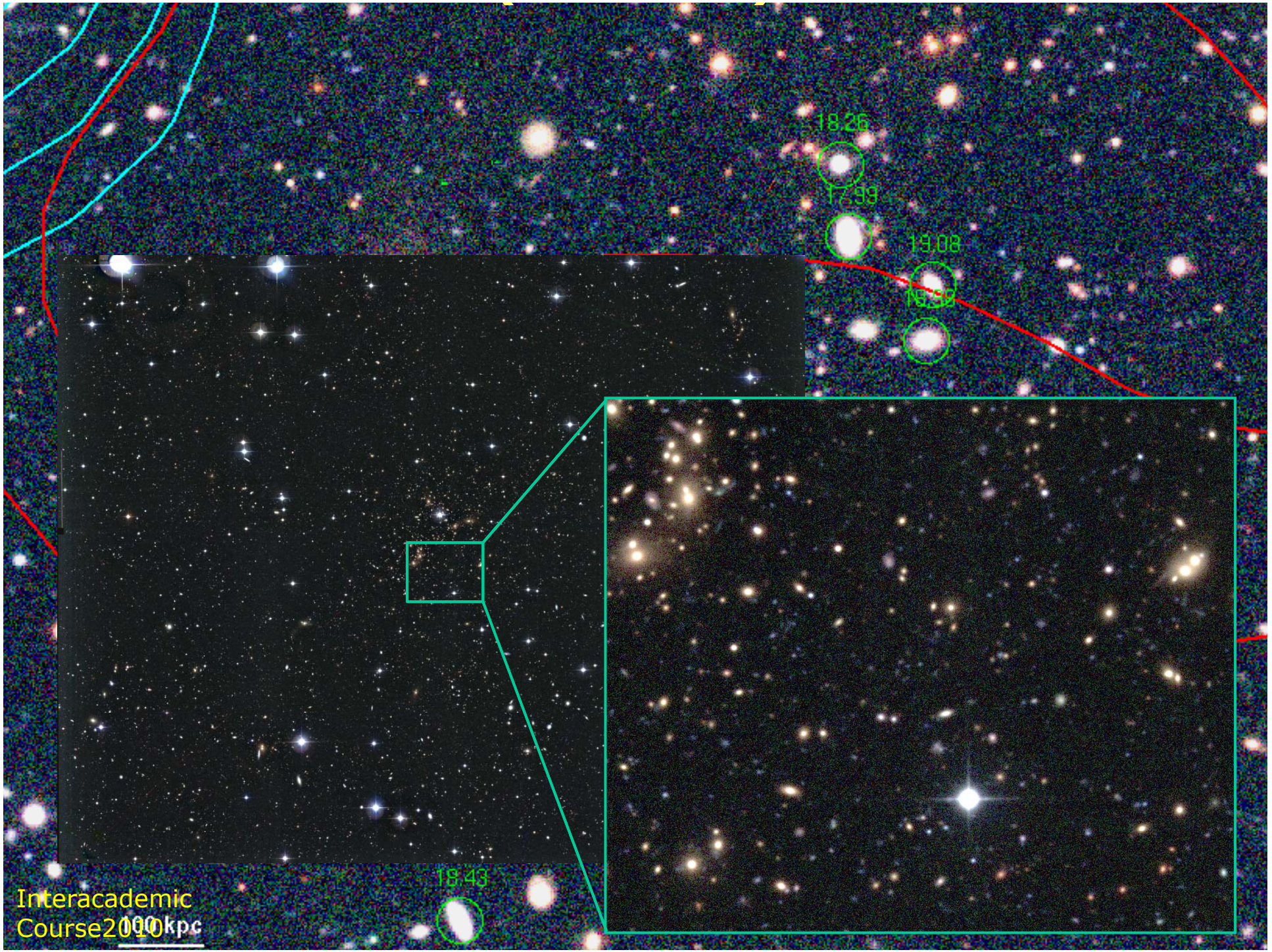
Fenemenology ←understanding→ theory

Observations - interpretation

Primary relations - secondary

N parameters N datapoints

Example Morphological transitions


18.26
17.99
19.08
18.43
Interacademic
Course2010
1000kpc

Braglia

Pierini

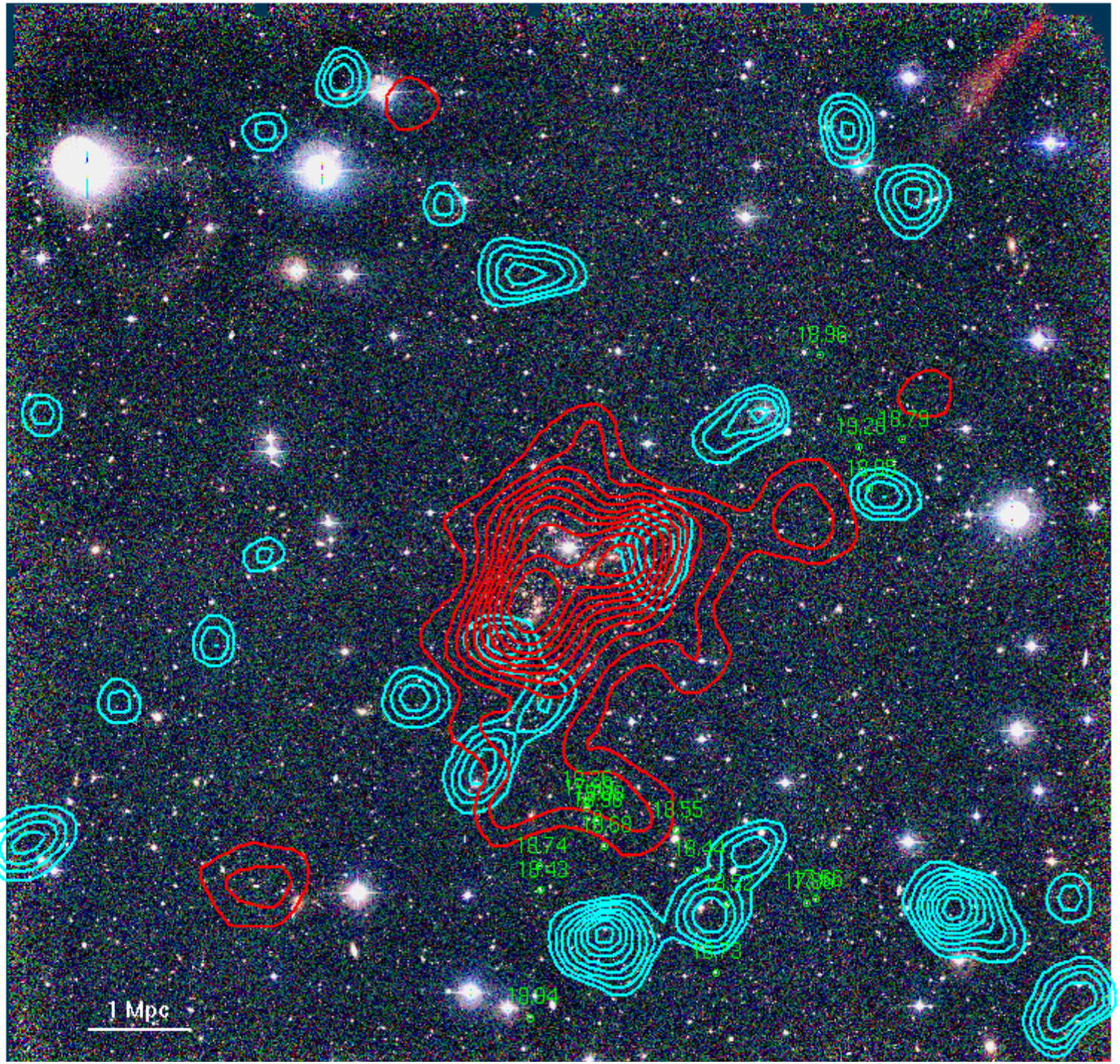
Bohringer

AA L 2007

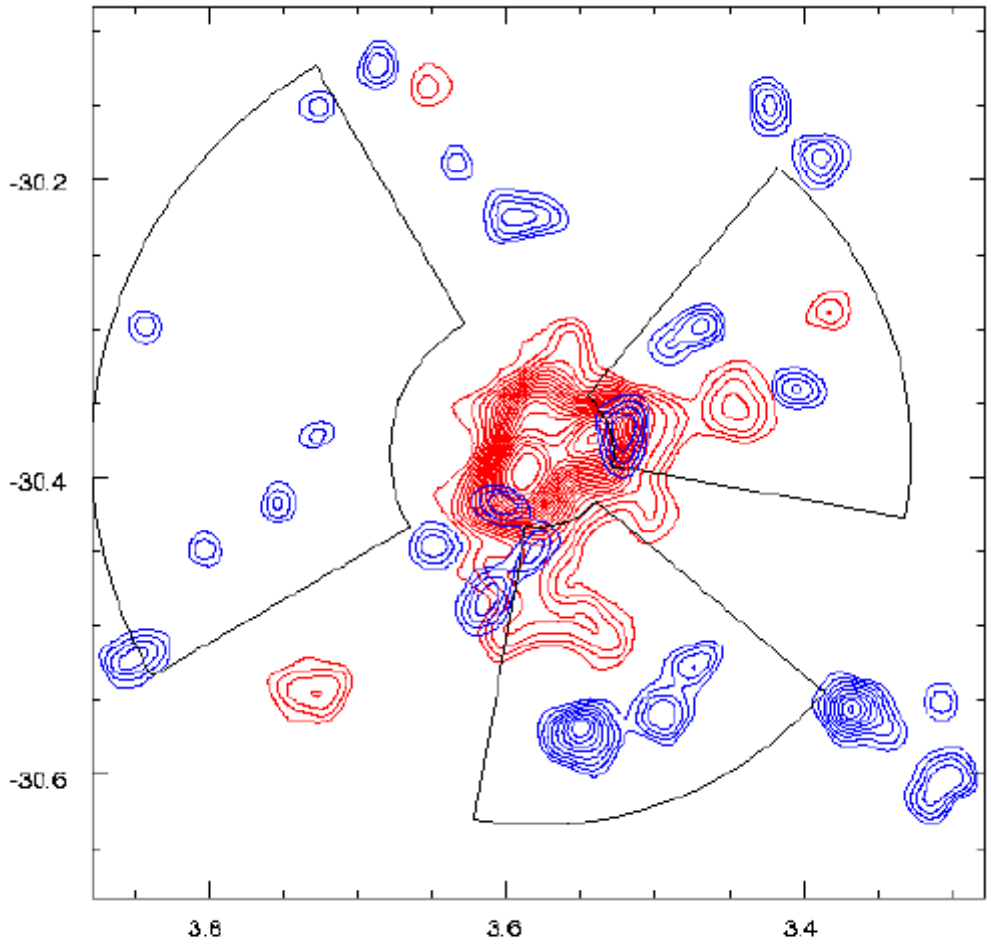
A2744

Z~0.3

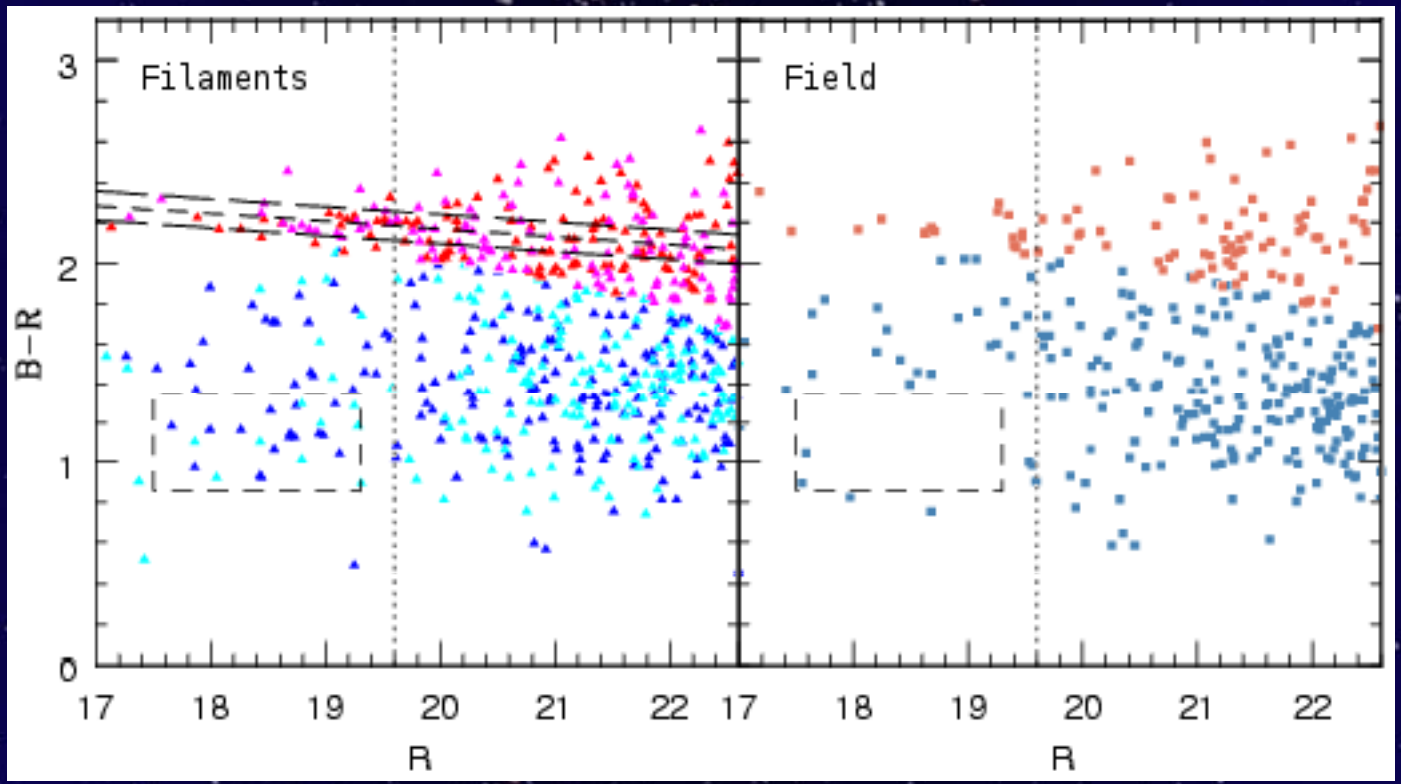
WFI@2.2m



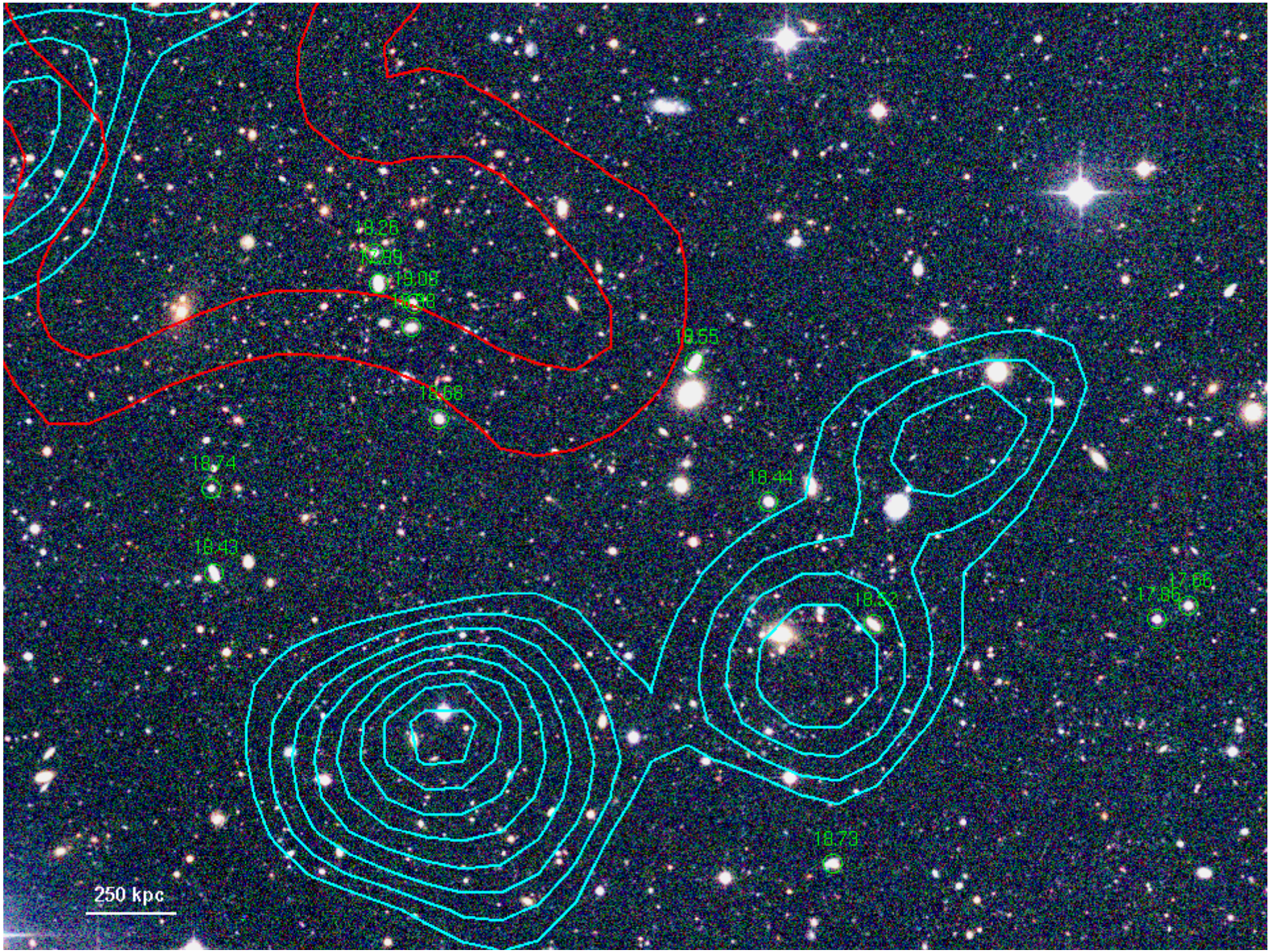
Interac
Course

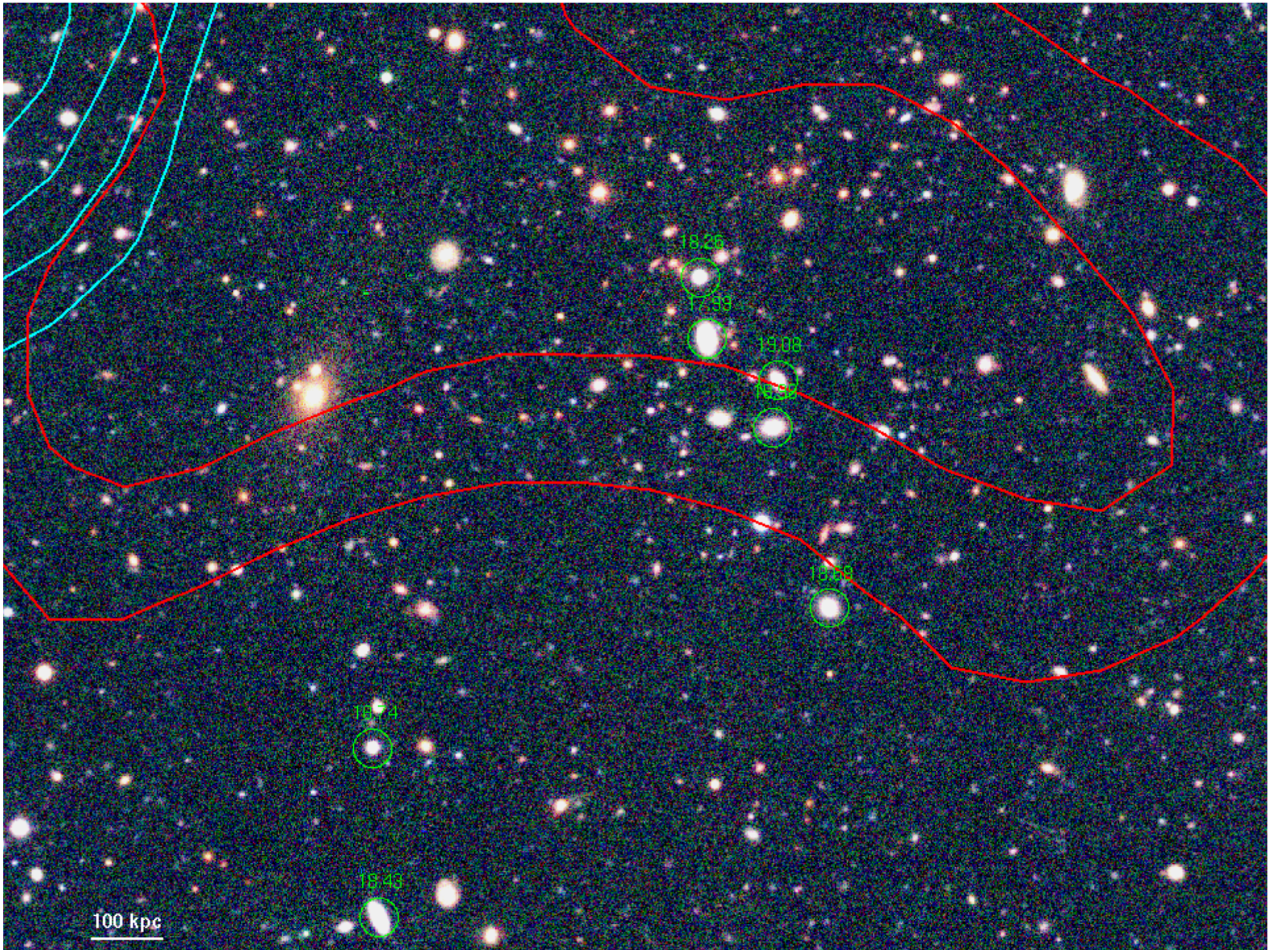


Interacademic Course2010



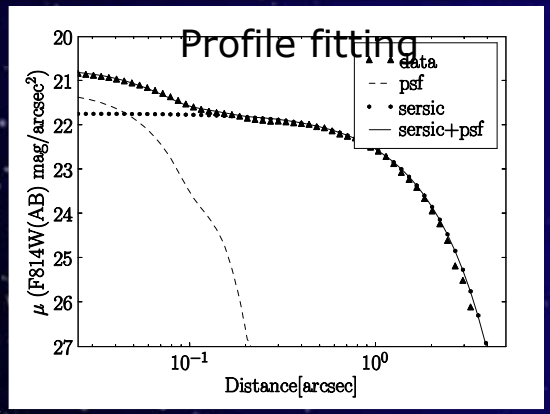
Interacademic Course2010

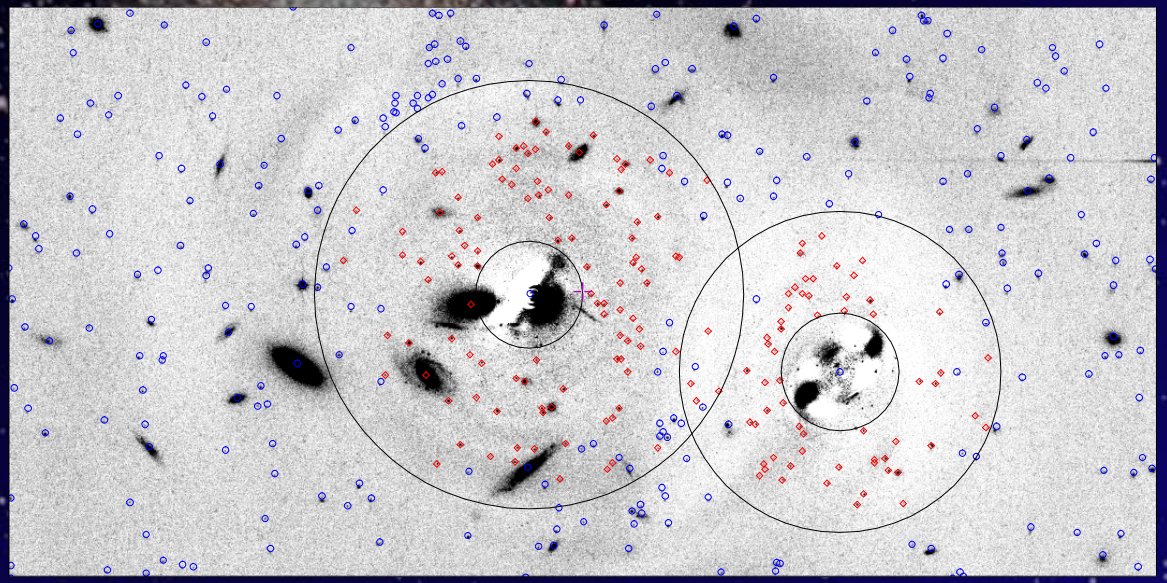

18.26
19.08
18.55
18.68
18.74
18.44
18.43
17.66
17.86
18.73
250 kpc


18.26
17.99
19.08
18.58
18.43
100 kpc

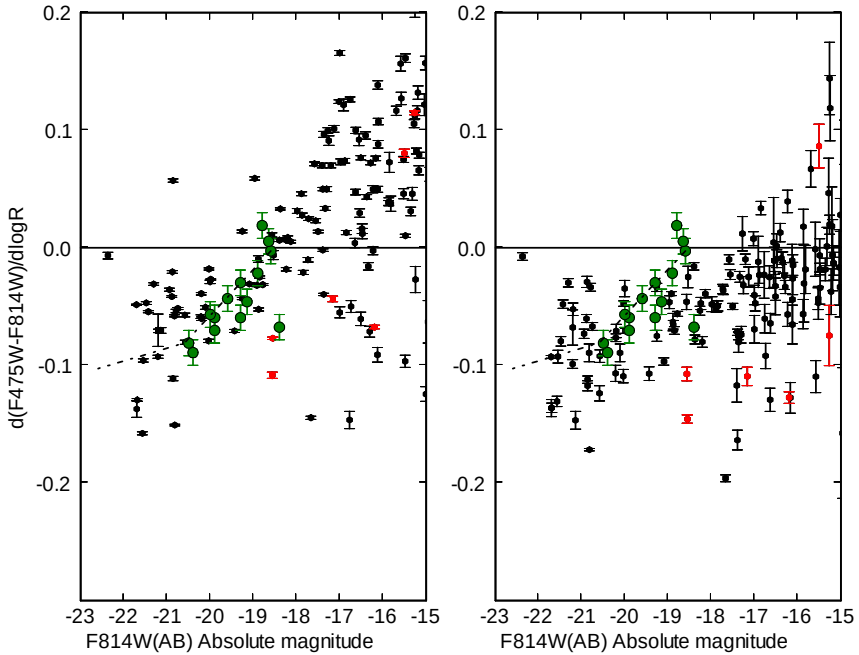
ACS@HST Coma LS

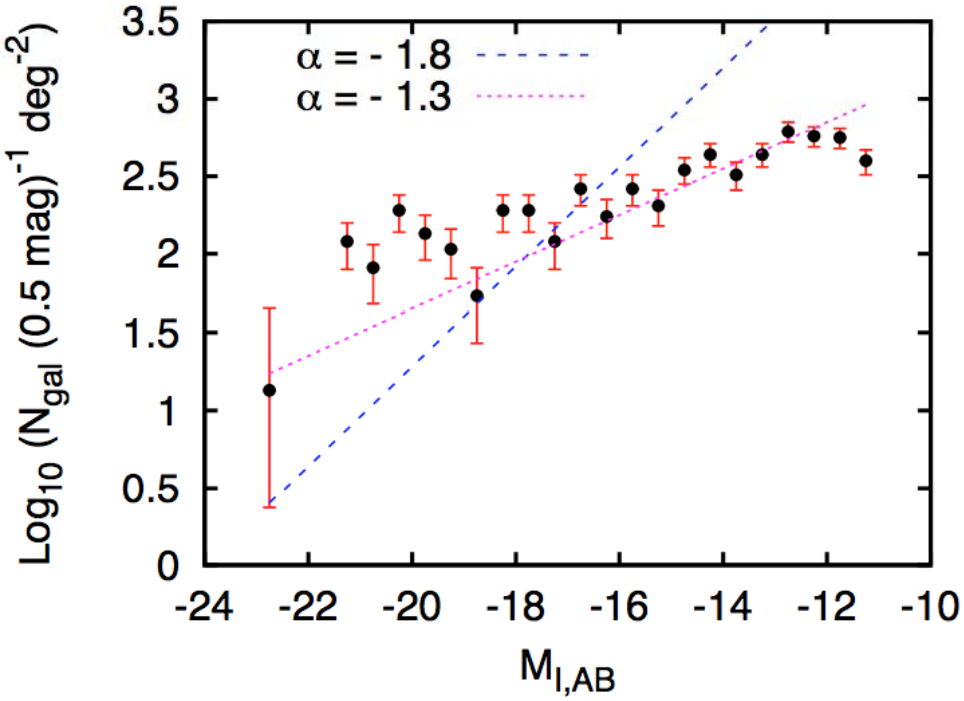
- AstroWise enables COMA LS team
- 10,000 objects via galphot/ galfit parallel processing
  1. Photometric catalog full survey 70,000
  2. Structural parameters catalog full survey





Carter et al. 2008; Hammer et al, 2010 submitted; Hoyos, et al., 2010, in prep



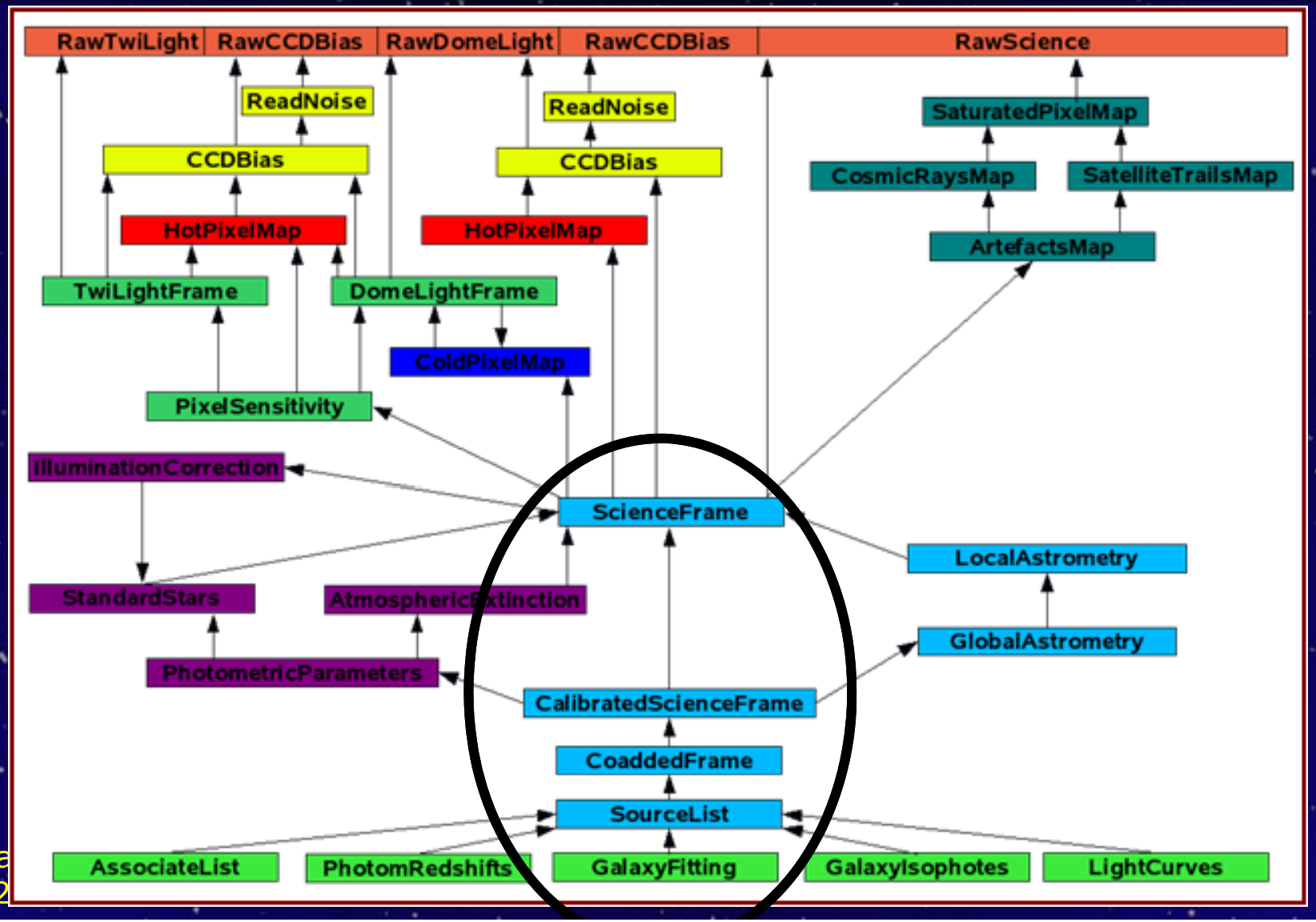


PhD Den Brok 2010

- Extreme faint-end luminosity function Slope $\alpha \approx -1.2$
- Much shallower than Cold dark Matter prediction $\alpha \approx -1.8$
- clean radial colour gradients are much smaller
- Hint of extensive Intracluster globular cluster population

Interacademic Course2010

TARGET diagram



N params N data back to basics

- Joins – links
- ++ Inheritance – dependencies

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

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*')
```

Example 5LS

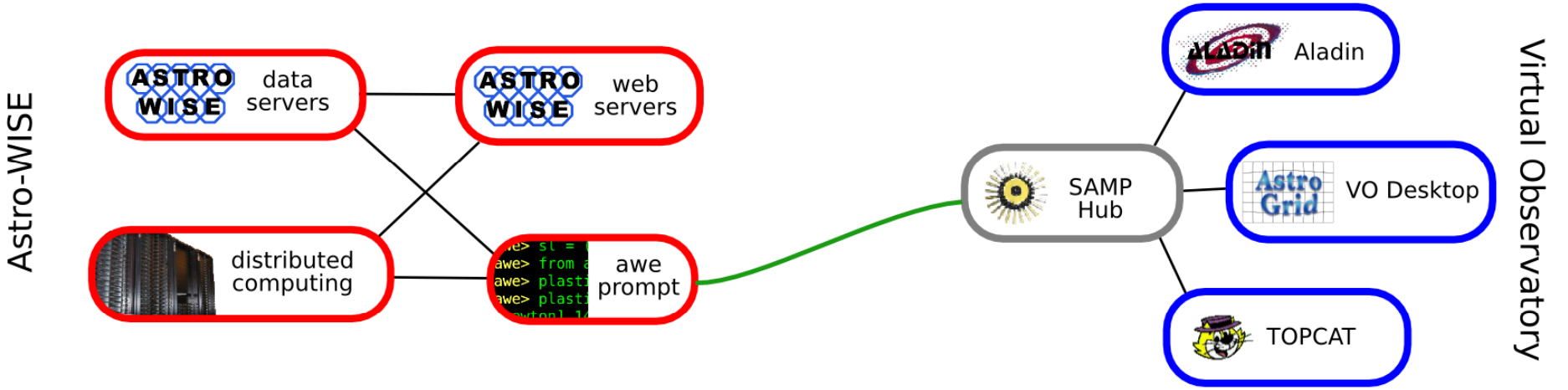
```
#Find ScienceFrames for a ccd named ccd53 and filter

    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)
```

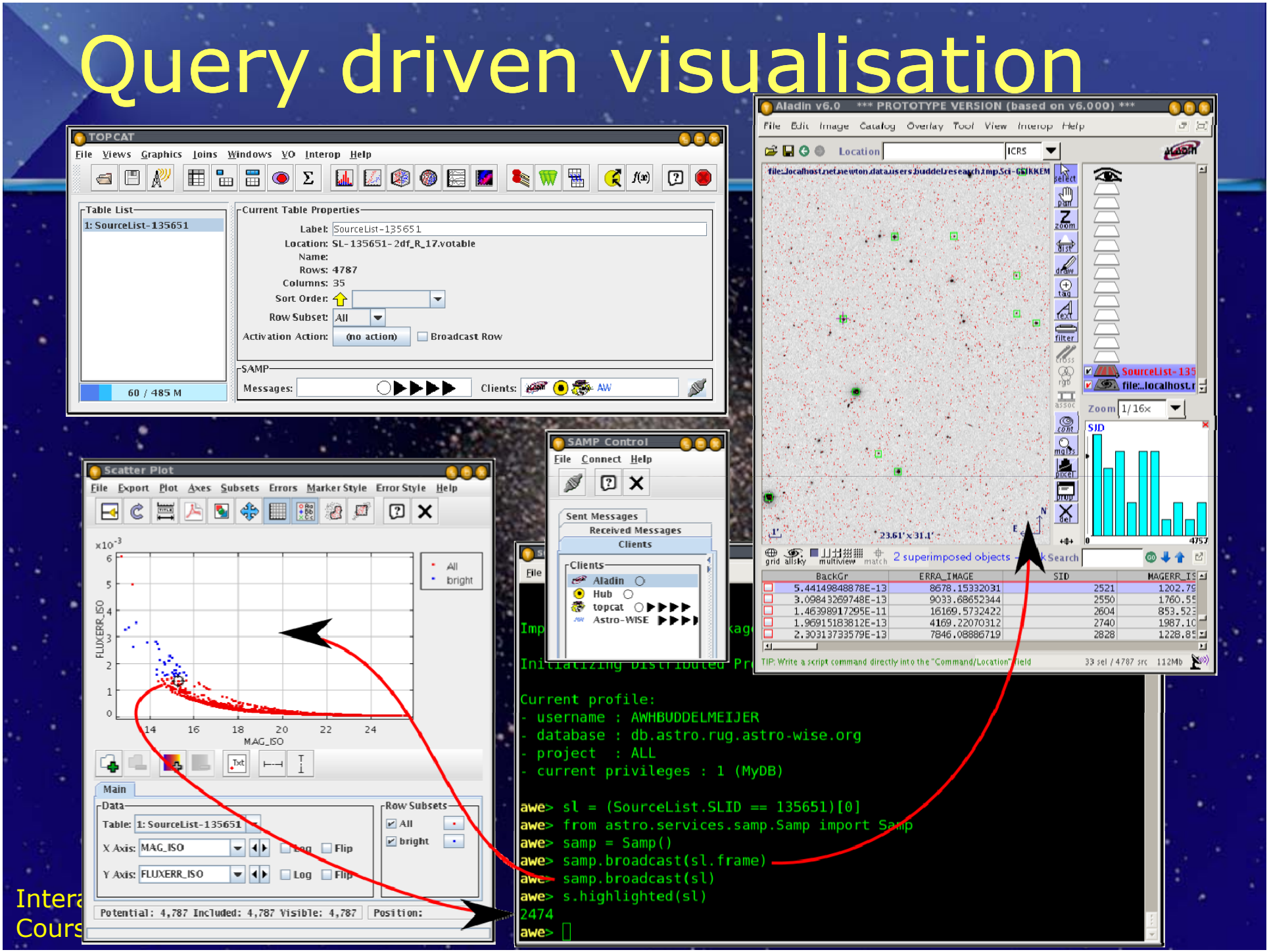
Query driven visualization



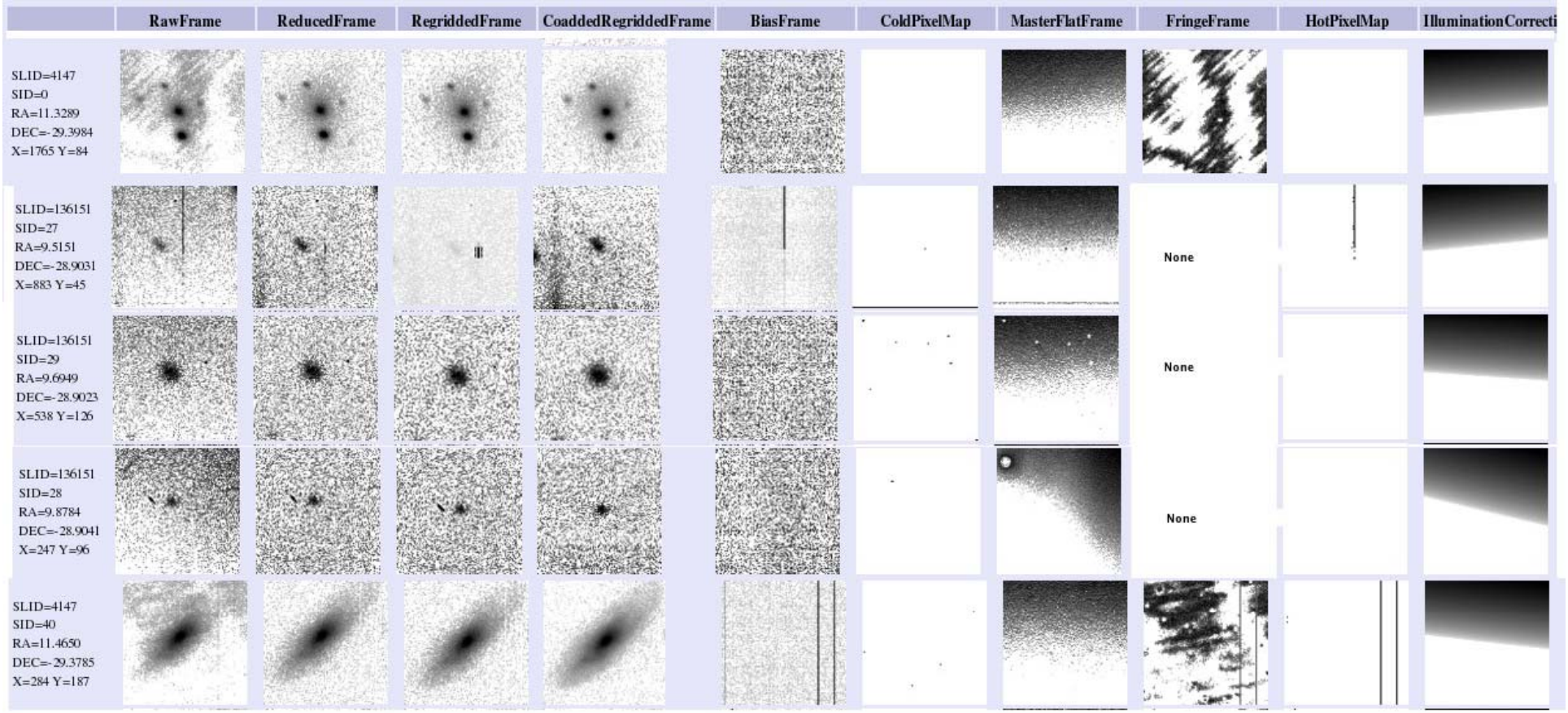
Ph D Buddelmeijer

Interacademic Course2010

Query driven visualisation



Extreme data lineage

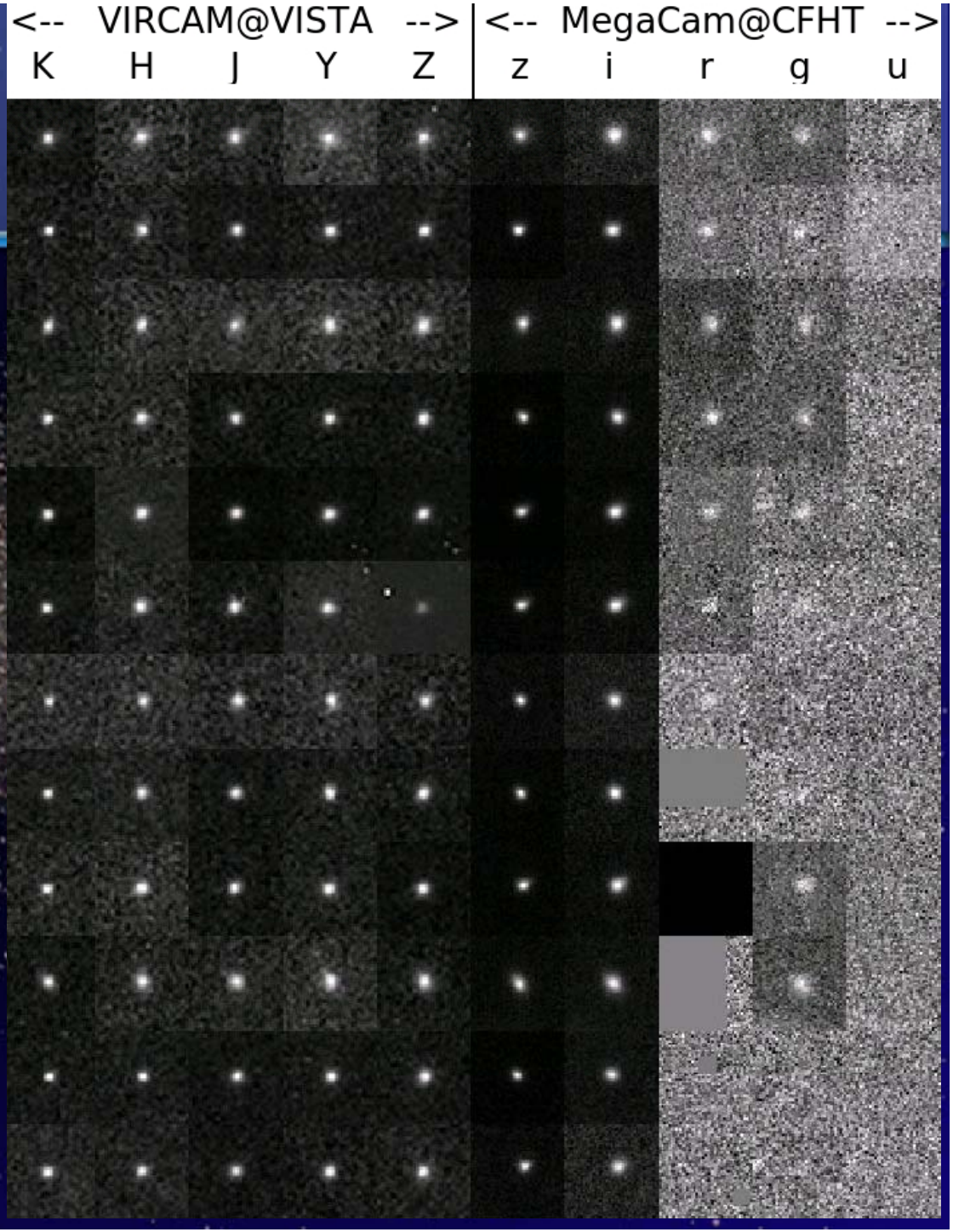


Comp. science journal PhD Mwebaze

Interacademic Course2010

Drop-outs

PhD Bout



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

time

Everything changes in time

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

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 DBView CalTS Process

no previous comments

DBname: swgverdoes 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	She-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

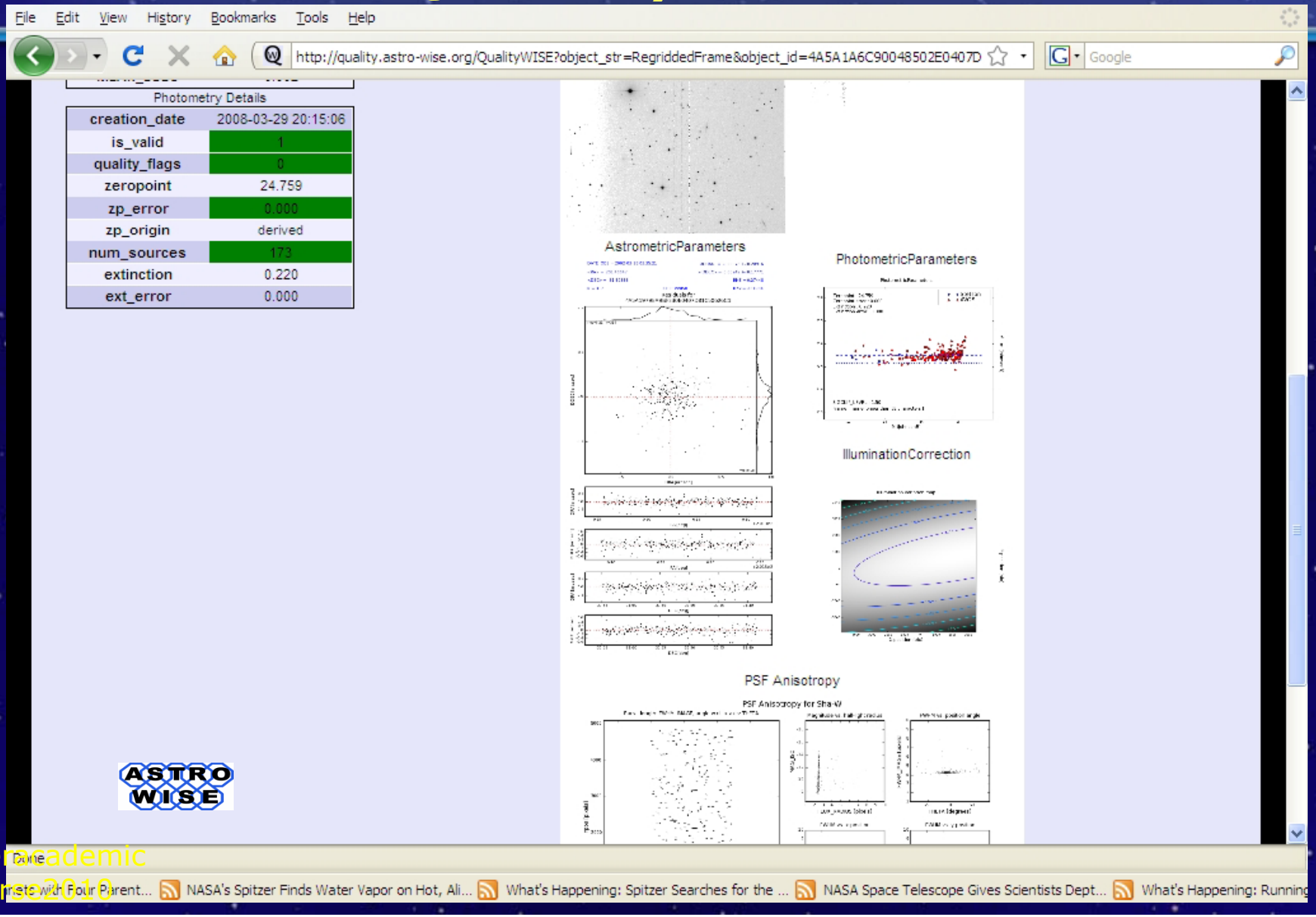
Done

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

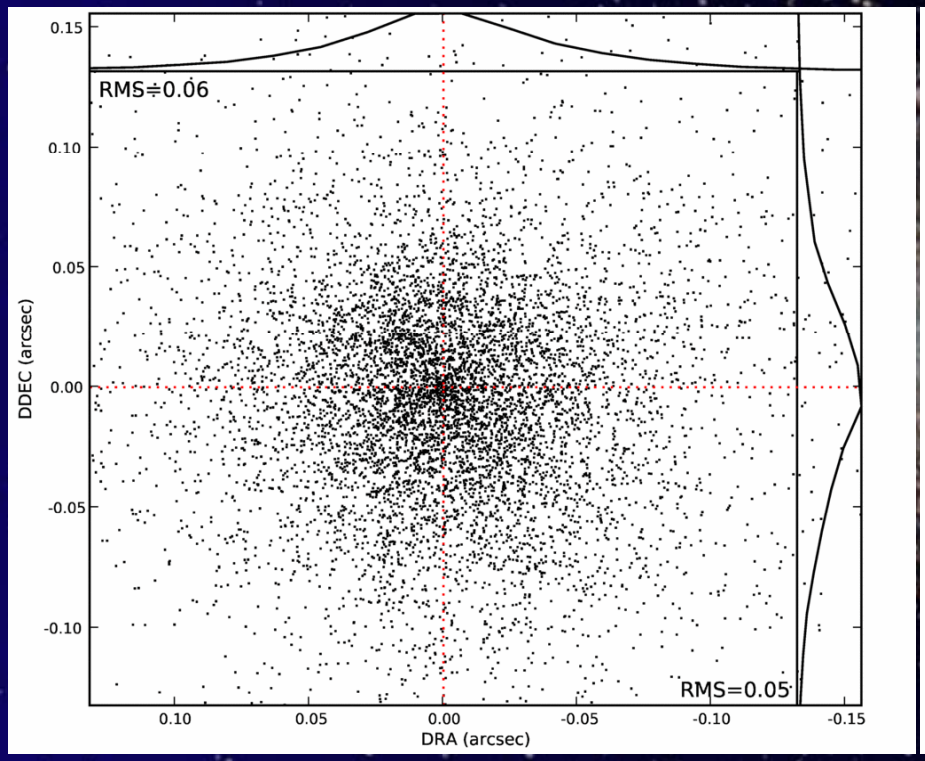
Interacademic Course2010

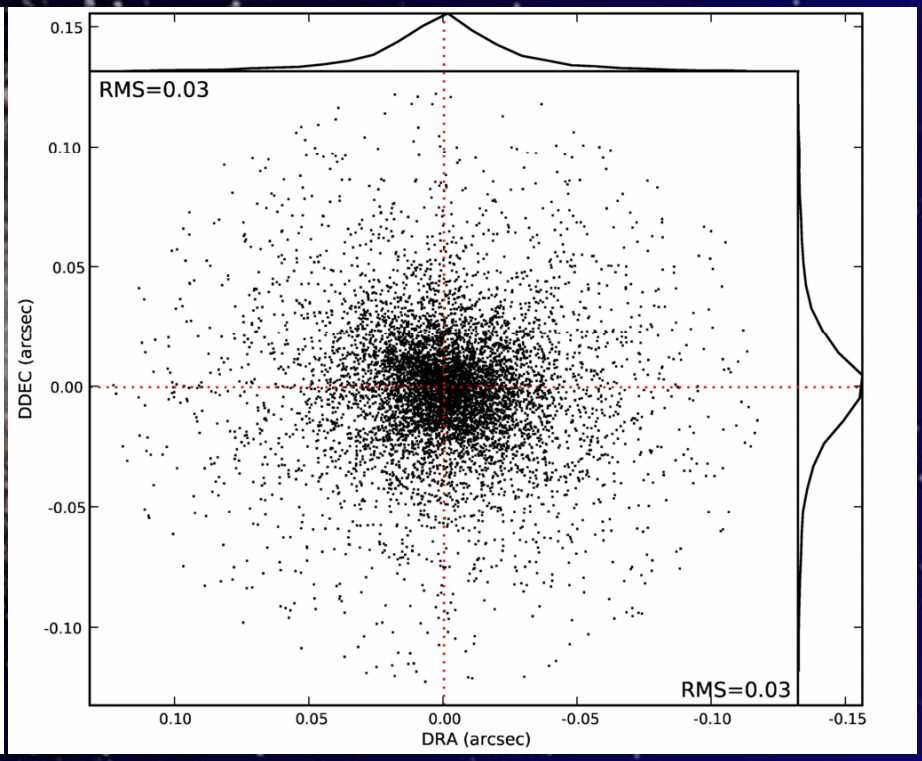
Post Doc Verdoes Kleijn

Quality view



Global Astrometry
reproducibility





Local solution

Global solution

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

Post Doc Mc Farland

future

- Information systems

TarGet

98 99 100 101 102 103 104

het probleem
Sensor netwerken-data centric

SENSOREN

Streaming processing

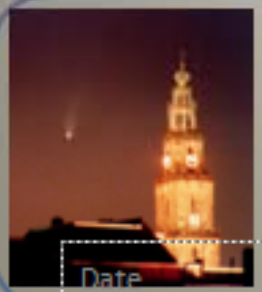
stor age

off-line processing

DB

web

TarGet

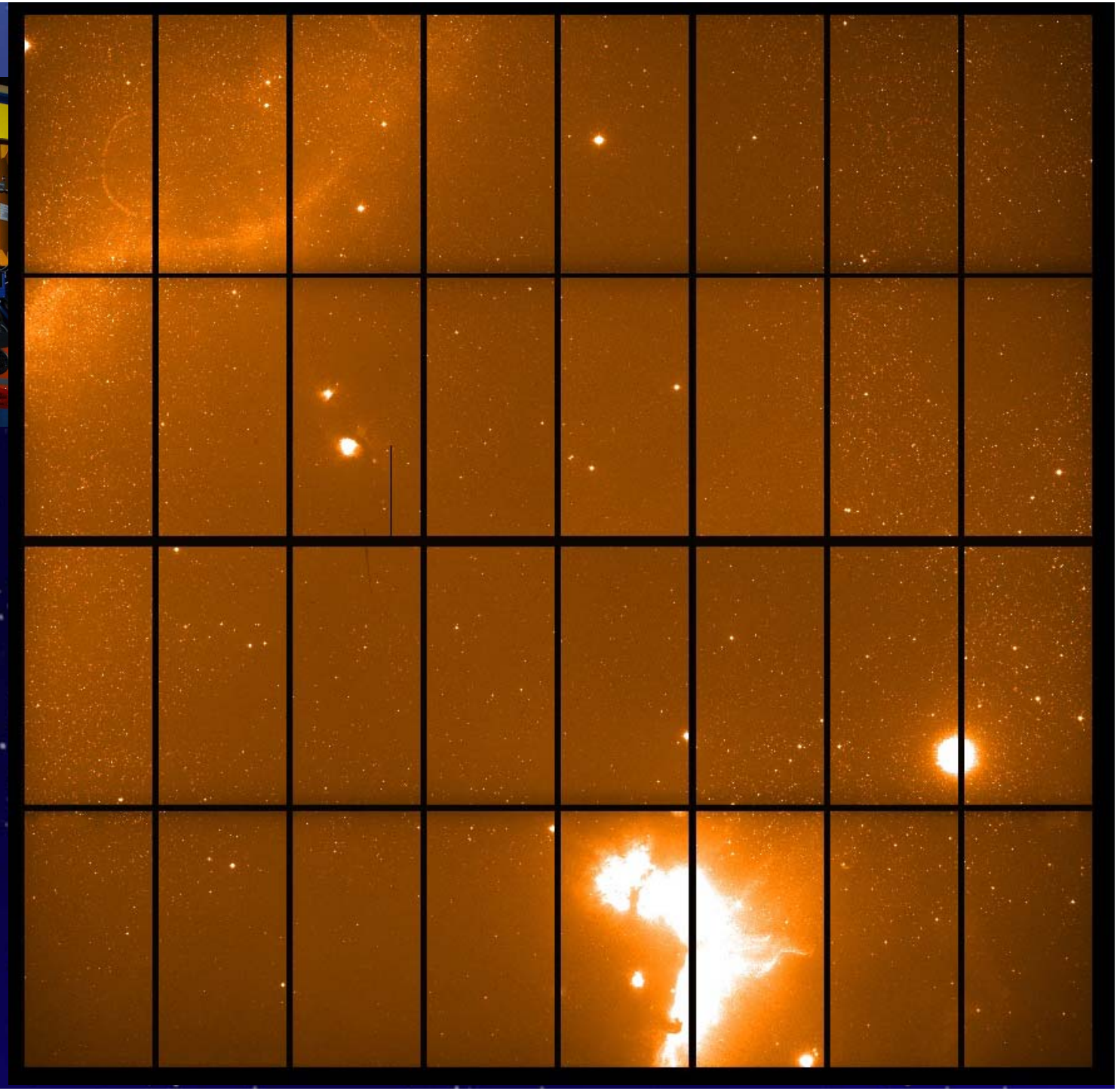


10 Petabyte =
10.000 Terabyte =
10.000.000 Gigabyte =
30 Martini torens DVD's

- 10+ Petabyte
- gedistribueerd

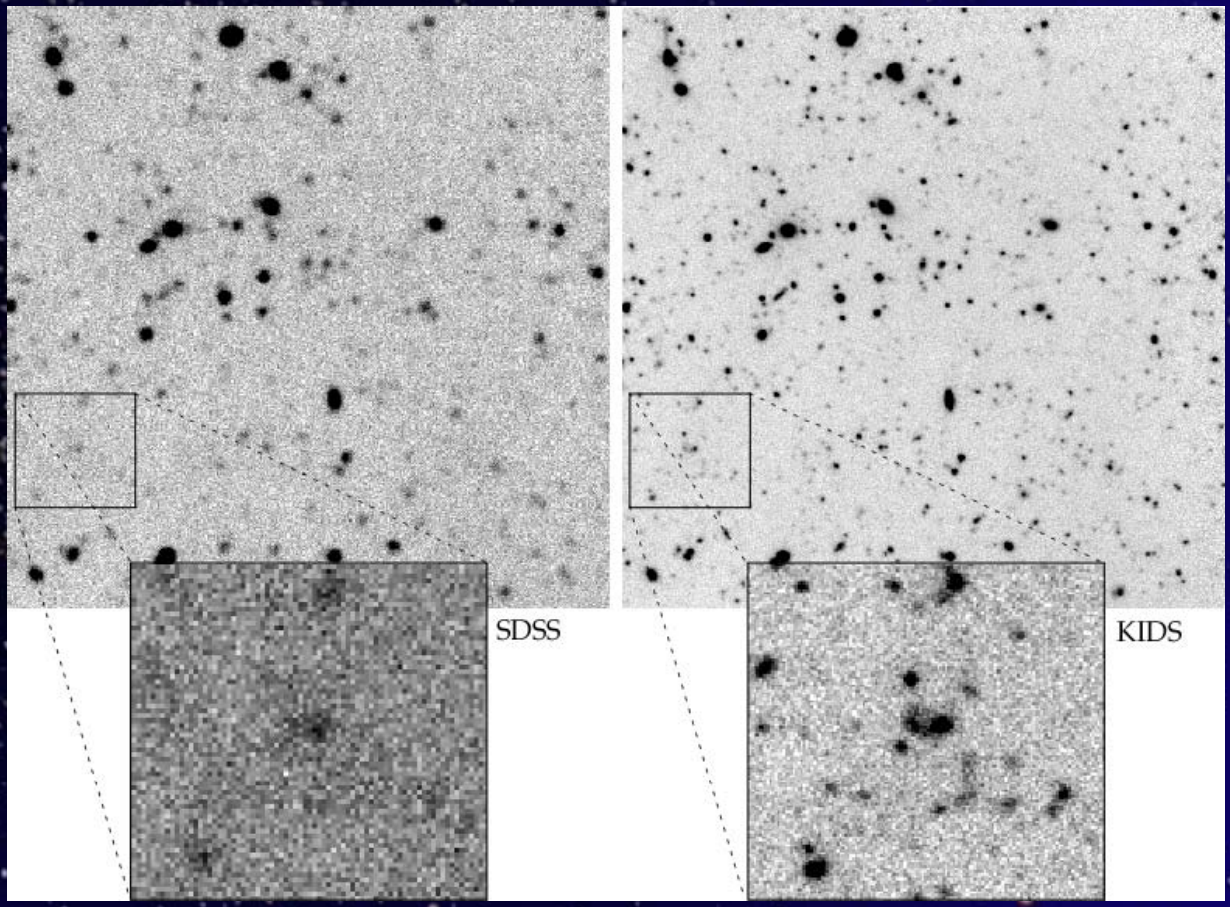
OmegaCAM

16k x 16k
256Mpix
0.2” pixels
1 x 1 degree
2.6 m VST



KIDS vs. SDSS, CFHTLS

SDSS	CFHTLS
6 x area	1/9th area
2 mag shallower	1 mag deeper
2x worse seeing	~same



(M.Neeser)

KIDS – VIKING

250 nights

440 nights

VIKING

VISTA

4m telescope

0.6 sq.deg. InfraRed camera

16 2kx2k detectors

0.35” pixels

KIDS

VST

2.6m telescope

1 sq.deg. optical cameraOmegaCAM

32 2kx4k detectors

0.21” pixels

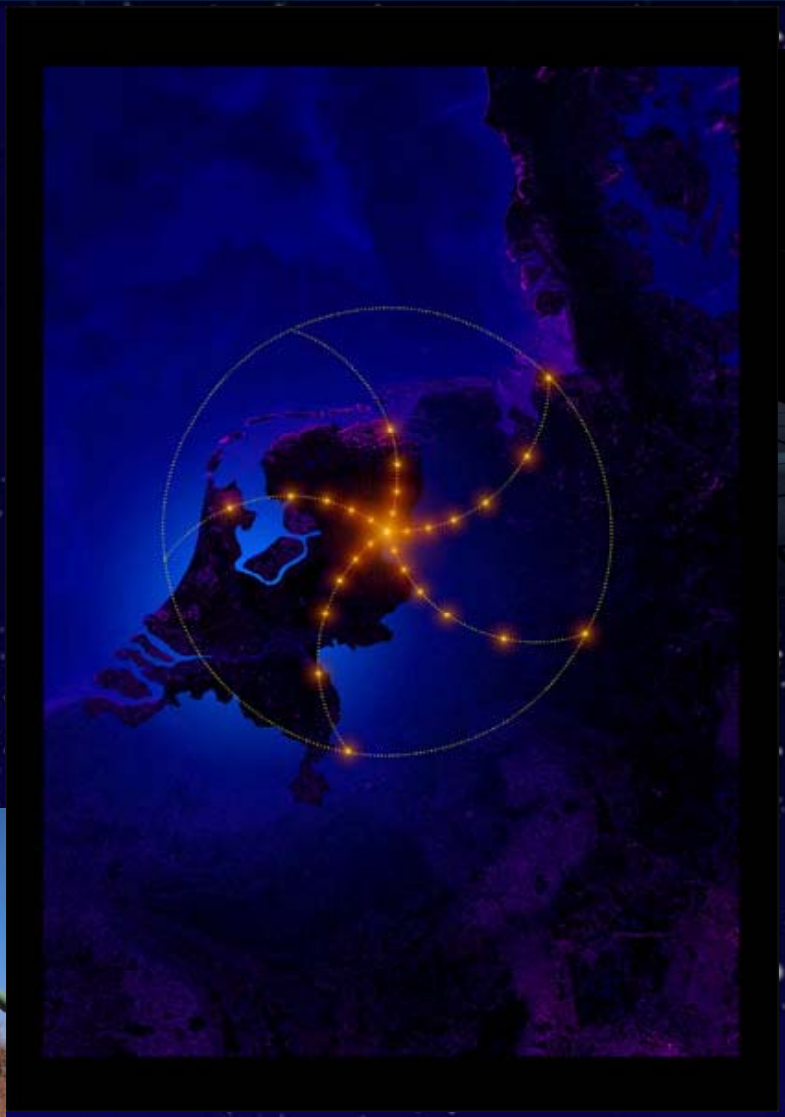
Paranal Monthly Data Rates 2007 statistics

All Current Paranal Instruments 4% (433.2 GB)

OmegaCAM 24% (~2500 GB)

VIRCAM 72% (~7500 GB)

Astro-WISE – LOFAR LTA



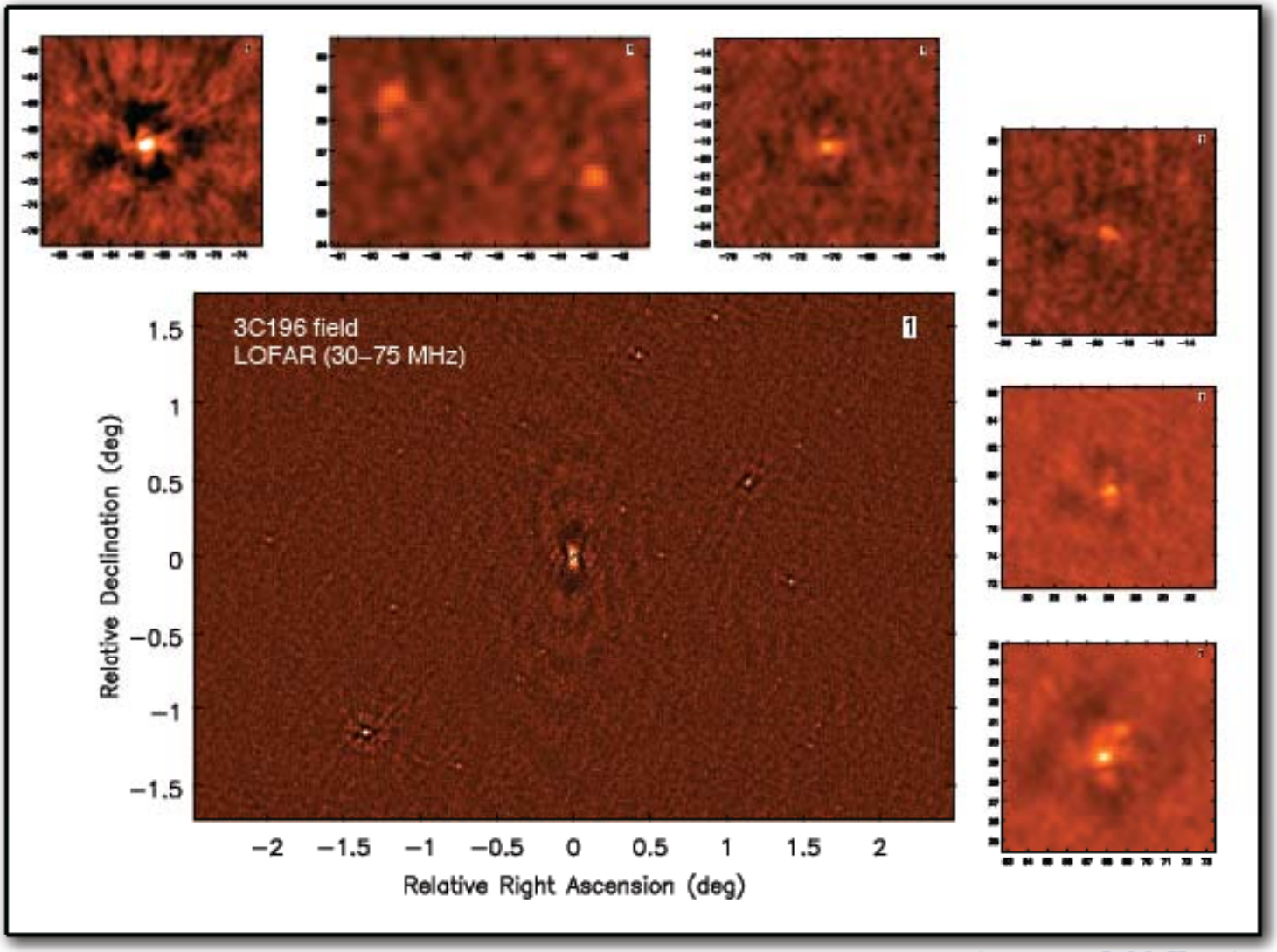
IBM- Blue Gene/P

5 Petabyte/yr

- EoR
- Survey


LOFAR Superterp
Interacademic
Course2010

Pipeline Processing

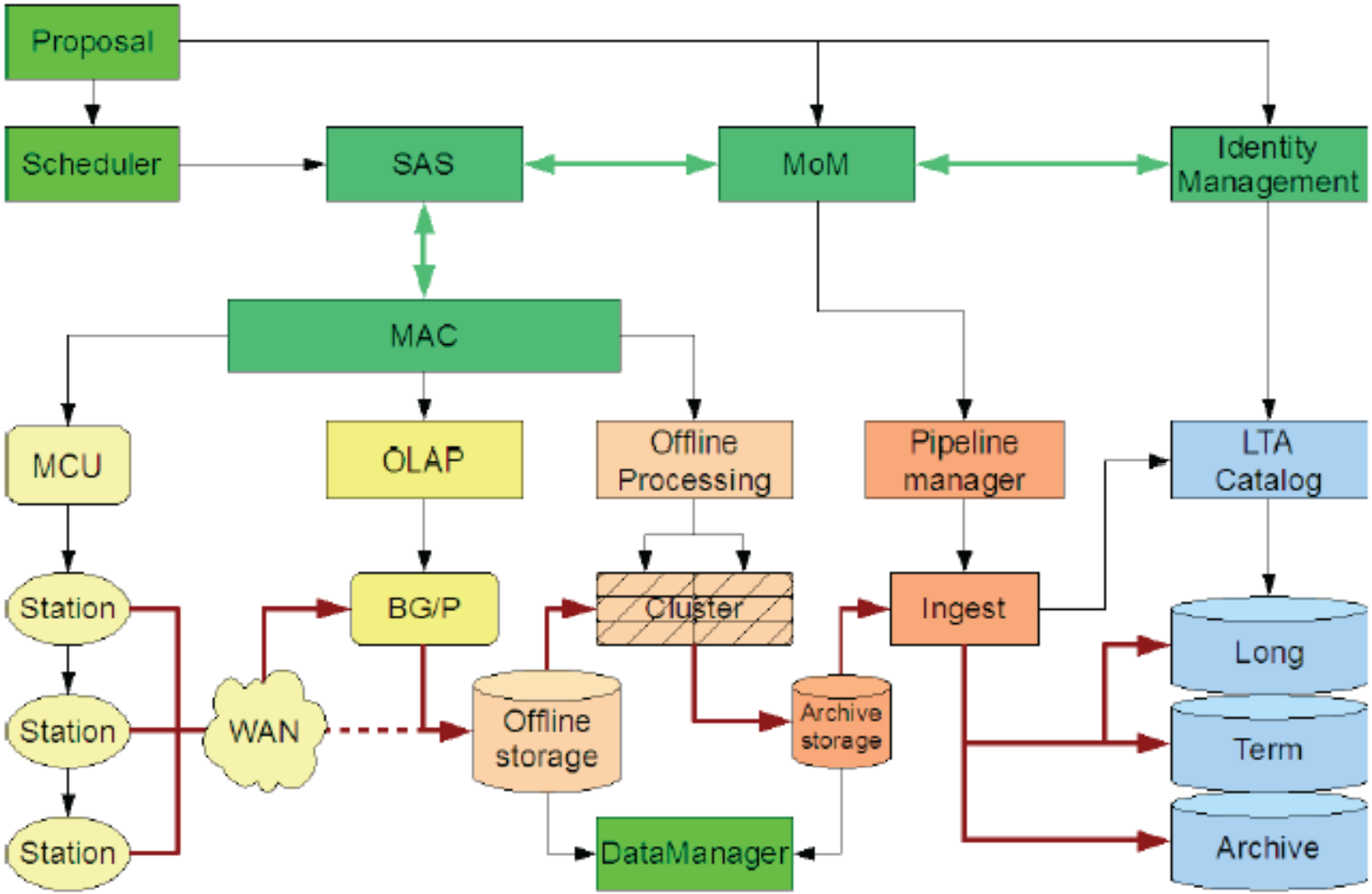


(courtesy J. McKean)

Fully automated processing, 72 sub-bands, ~4 hrs
Included DPPP, additional flagging pass, BBS, solution flagging, imaging

Interacademic Course 2010

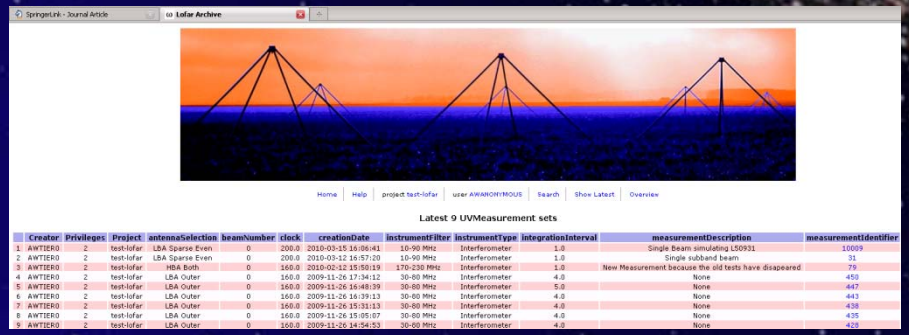
Long Term Archive

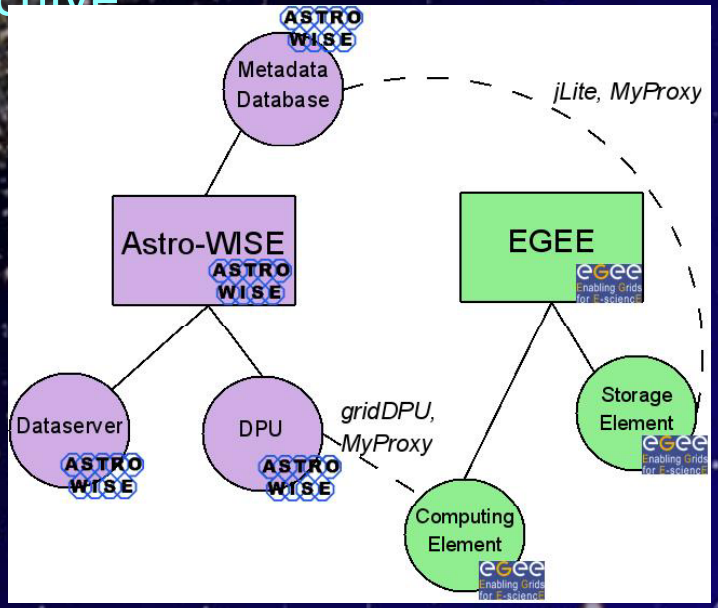


(courtesy H. Holtes)

Lofar Long Term Archive

- Astro-WISE –> EGEE full integration
  - Astro-WISE stores on EGEE
  - launches Astro-WISE applications on EGEE
  - Incorporated 3rd party software gLite,MyProxy
- Designed+implemented LOFAR LTA
  - Ready for >10 Petabyte
  - Next step: integration pipeline+archive

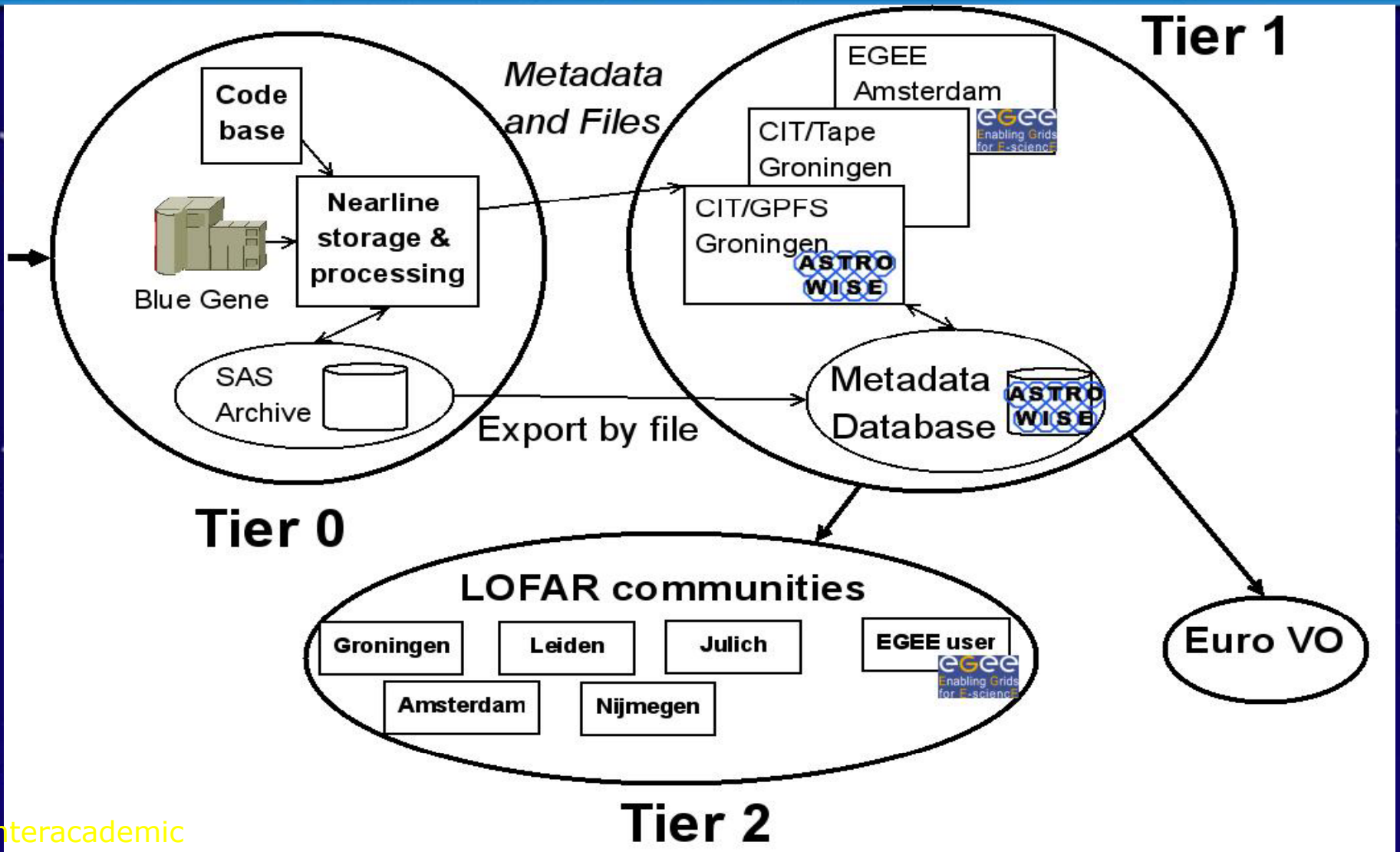




Interacademic Course2010

Valentijn & Belikov 2009; Begeman et al 2009+2010

Lofar - design



Estimated Data Volumes

	60 MHz	150 MHz
Total # fields (2 pi steradian)	609	3346
Total observing time (100% eff., using 4 beams)	456.75 hr	836.5 hr
Total # sources	Tbd	Tbd
Total uv data size	466 Tbyte	2.27 Pbyte
Total post DP^3 uv data size	~ 2.9 Tbyte	~ 7.1 Tbyte
Total image data size	Tbd	Tbd

(R. Nijboer, March 2009)

Current archive resources

- Currently using ~50 TBytes
- 300 TBytes available @ SARA in Amsterdam
- 60 TBytes @ Astrowise server in Groningen
- 500 TBytes (to 1 PBytes) via TARGET in Groningen
- 1 PBytes tape storage @ Jülich
- Additional 2-3 PBytes to be added @ BiGGRID in Amsterdam

- Lofar 2010 www.lofar.org
- ALMA 2013 www.eso.org/sci/facilities/alma
- EUCLID 2018
- LSST 30Tb/night www.lsst.org
- SKA 2022 www.skatelescope.org

- CLOUD – computing
- Web 2.0

- Crowd sourcing
  - Galaxy zoo
  - annotation

AstroInformatics

- www.astroinformatics2010.org

Sunday May 09 2010 13:06 PST — 38 days until the conference

AstroInformatics 2010
- Introduction
- 42 Participants
- Conference Agenda
- Invited Speakers
- Proceedings Info
- Registration
- Pay by CreditCard

More Information
- Accomodation
- Getting to Caltech
- Local Information
- Conference Poster
- Conference Photos
- Organizing Committee

Contact the LOC

AstroInformatics 2010

Caltech, June 16-19 (Wed-Sat), 2010
Cahill Center for Astronomy & Astrophysics, Caltech, Pasadena, CA, USA

Conference motivation and goals:

Science is being fundamentally transformed and empowered by computation and information technology, and it turn, it stimulates further technological developments. Astronomy, like many other fields, is becoming exponentially data-rich, and the tasks of data management, data exploration, and knowledge discovery become central to our research enterprise, bringing along many technical and methodological challenges. Information technology also provides the stage where we collaborate and interact, and publish, preserve, and disseminate knowledge.

The general philosophy behind the conference is to be future-oriented, and essentially define the emerging discipline of AstroInformatics. One way to think about it is, what is next for, and beyond the VO, scientifically, technologically, institutionally? AstroInformatics can be viewed as a broader intellectual, organizational, and funding environment, within which VOs serve as particular institutions and provide some fundamental functionalities and infrastructure.

There will be a modest number of invited review talks to serve as a basis for the discussion, and a lot of discussion, some of it led by panels. Contributed papers will be accomodated as posters.

Interacademic Course2010

ADASS

Astronomical Data Analysis Software & Systems

The **Astronomical Data Analysis Software and Systems (ADASS)** conference is held each year at a different hosting astronomical institution. The conference provides a forum for scientists and programmers concerned with algorithms, software and software systems employed in the acquisition, reduction, analysis, and dissemination of astronomical data. An important element of the program is to foster communication between developers and users with a range of expertise in the production and use of software and systems. The program consists of invited talks, contributed oral and display papers, tutorials, user group meetings and special interest group meetings (called BOFs).

ADASS XX

ADASS XX

- The 20th annual ADASS conference will be in Boston, Massachusetts, USA, 7-11 November 2010.

EURO-VO IVAO

Welcome to the Euro-VO AIDA Astronomical Infrastructure for Data Access

This is the web-based collaboration area of the Euro-VO Astronomical Infrastructure for Data Access project. This project is supported by EU in the framework of the FP7 eInfrastructure Scientific Research Repositories initiative (project RI2121104). It started on 1 February 2008, for a duration of 30 months.

Highlights

- Fifth Euro-VO Technology Forum : March 16 -18, 2010 in Heidelberg
- Second Community feedback Workshop : 25-28 January 2010, Strasbourg
- **The EuroVO-AIDA Project also realizes an update of the DCA Census of European astronomical Data Centres.**
- EURO-VO in the ICT results: e-Infrastructures give real boost to virtual observatories - issued on 8 october 2009.
- Fourth Euro-VO Technology Forum: 22 - 24 September 2009, Trieste
- Second EuroVO-AIDA Research Initiative. Deadline 15 July 2009.
- Data Centre Workshop on how to publish data in the VO 22-26 June 2009, ESAC, Villafranca del Castillo
- EuroVO-AIDA Hands-on workshop : 30 March - 2 April 2009, ESO, Garching
- Third Euro-VO Technology Forum : 16- 18 March 2009, Strasbourg *Due to a strike warning on March 19, the meeting will be held March 16 -18*
- Full Harvestable EuroVO Registry of Resources has been released (13 Mar 2009)
- EuroVO-AIDA Workshop MultiWavelength astronomy and the Virtual Observatory : December 1-3 2008, ESAC, Villafranca del Castillo
- Second Euro-VO Technology Forum (with VO-TECH) : 29 September - 2 October 2008, Cambridge, UK
- First EuroVO-AIDA Research Initiative. Deadline 15 June 2008.
- First Euro-VO Technology Forum (with VO-TECH): March 17 - 19 2008, Strasbourg

The Reading Monk - Mozilla Firefox

Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help

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

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

TarGet

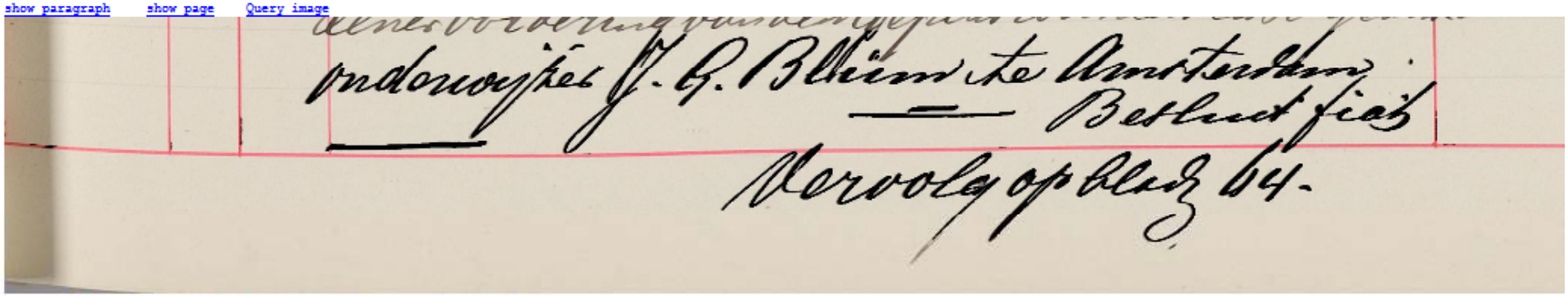
98 99 100 101 102 103 104

MONK

amsterdam Submit

There are 182 results. Showing results 1 - 10 for amsterdam

show paragraph show page Query image

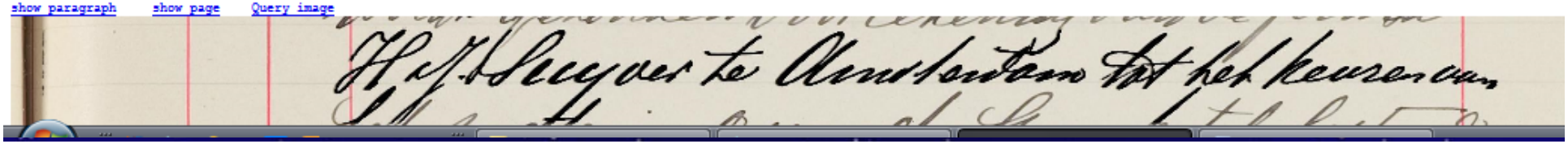


onderwijzer J.G. Bluem te Amsterdam. _-_ Besluit fiat ____ Vervolg op bladz 64.

show paragraph show page Query image

loniaal Etablissement te Amsterdam

show paragraph show page Query image

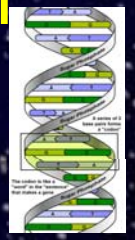


LifeLines
165.000 patients 30 yr

TarGet

98 99 100 101 102 103 104

Molecular



Physiology

Care

MOLGENIS

METABASE

ORACLE

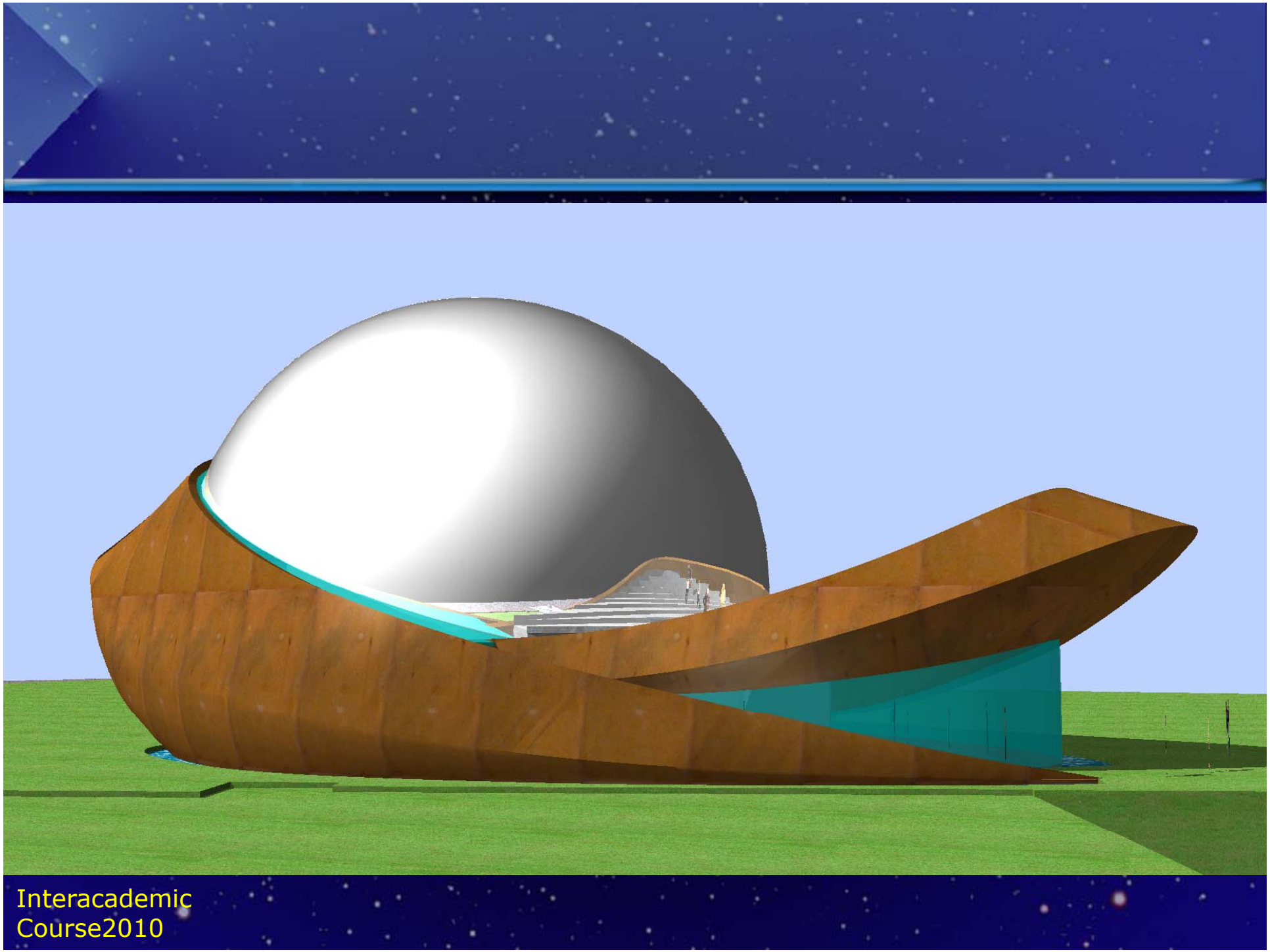
Healtcare Transaction Base
(HTB/HL7)

Integration

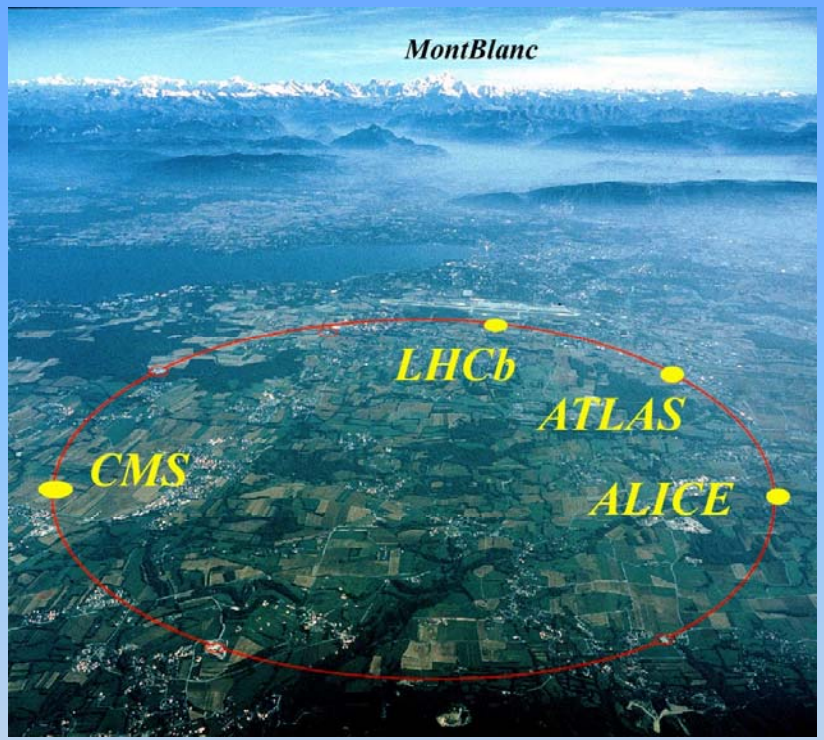
ASTRO WISE

Interacademic Course2010

IJkdijk


Interacademic
Course2010

LHC & EGEE



- 15 PB data per year
- grid computing

EGEE: ~250 sites, >45000 CPU

www.eu-egee.org

EGEE-III

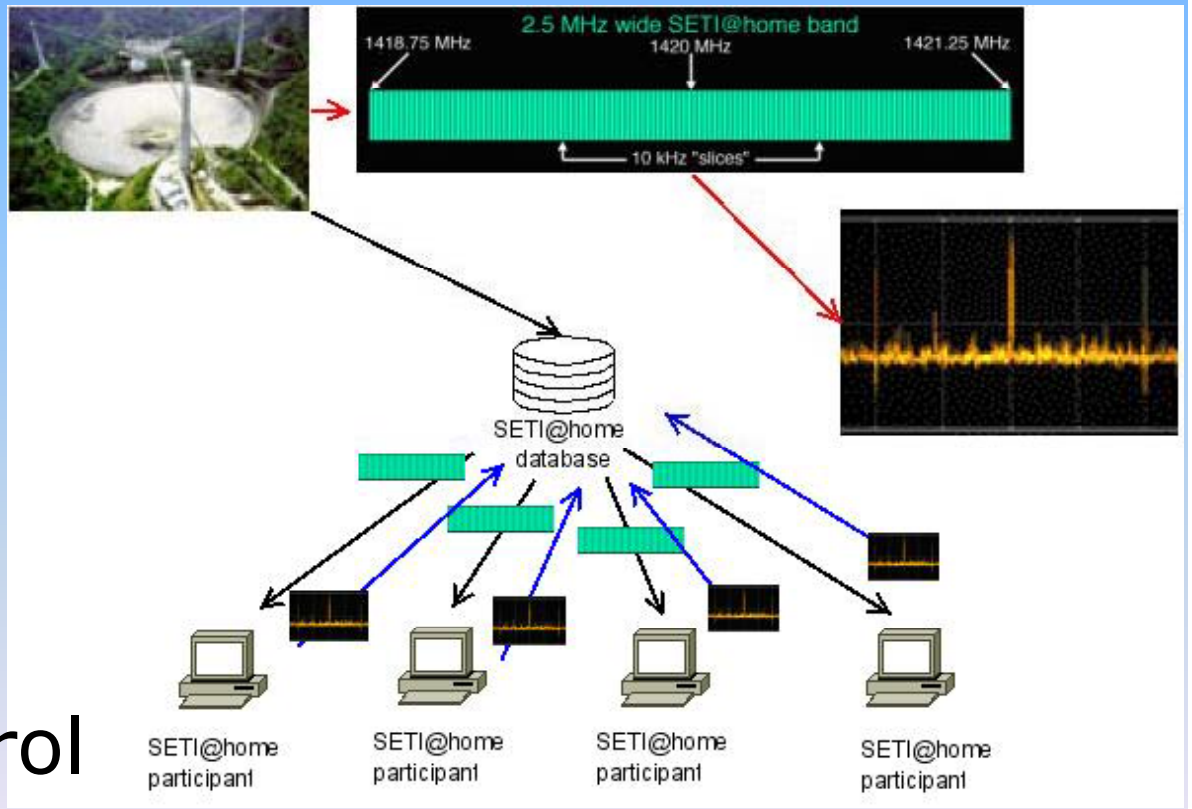
Enabling Grids for E-sciencE

- First scientific grid computing

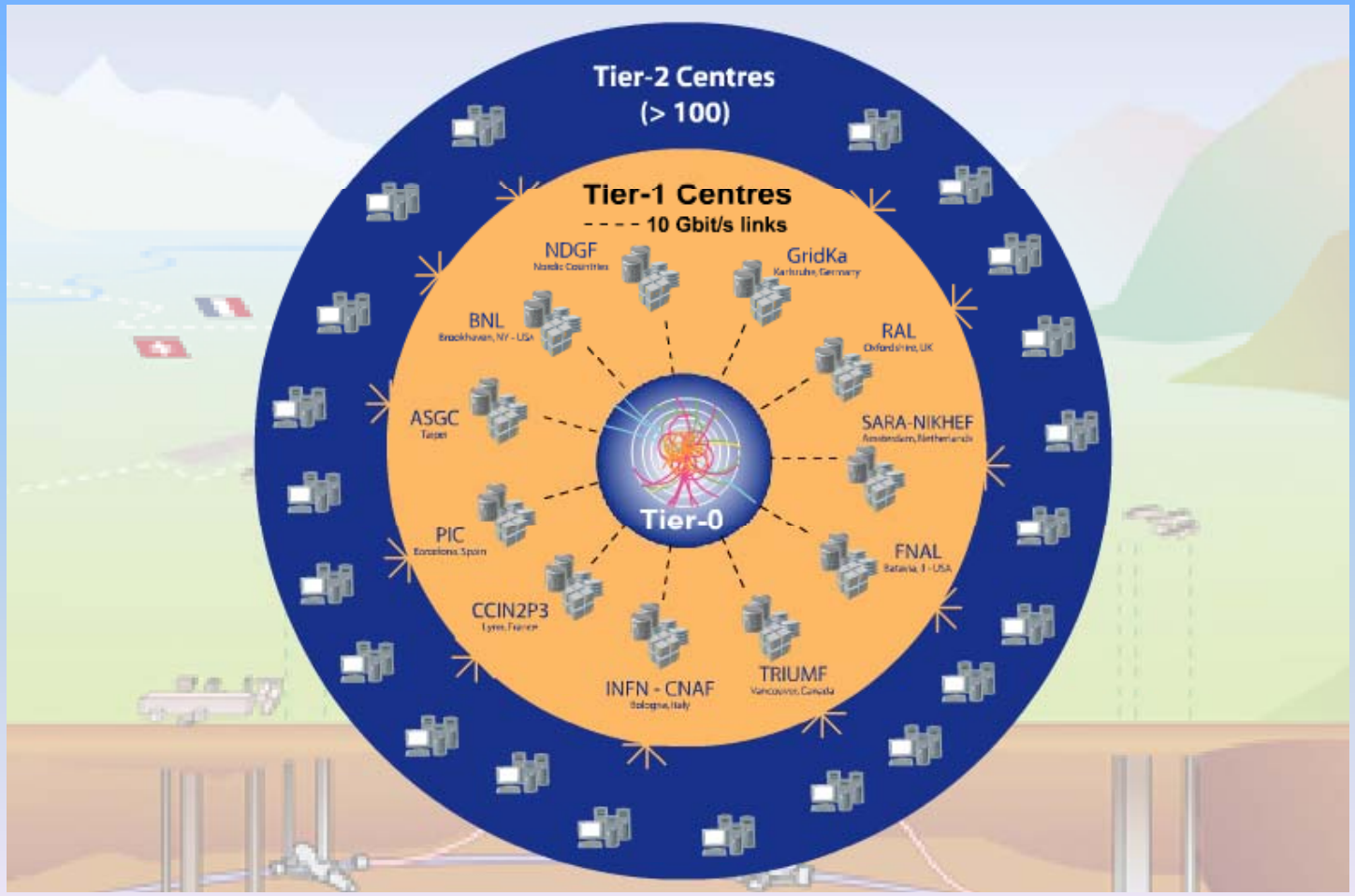
http://setiathome.berkeley.edu/

http://boinc.berkeley.edu

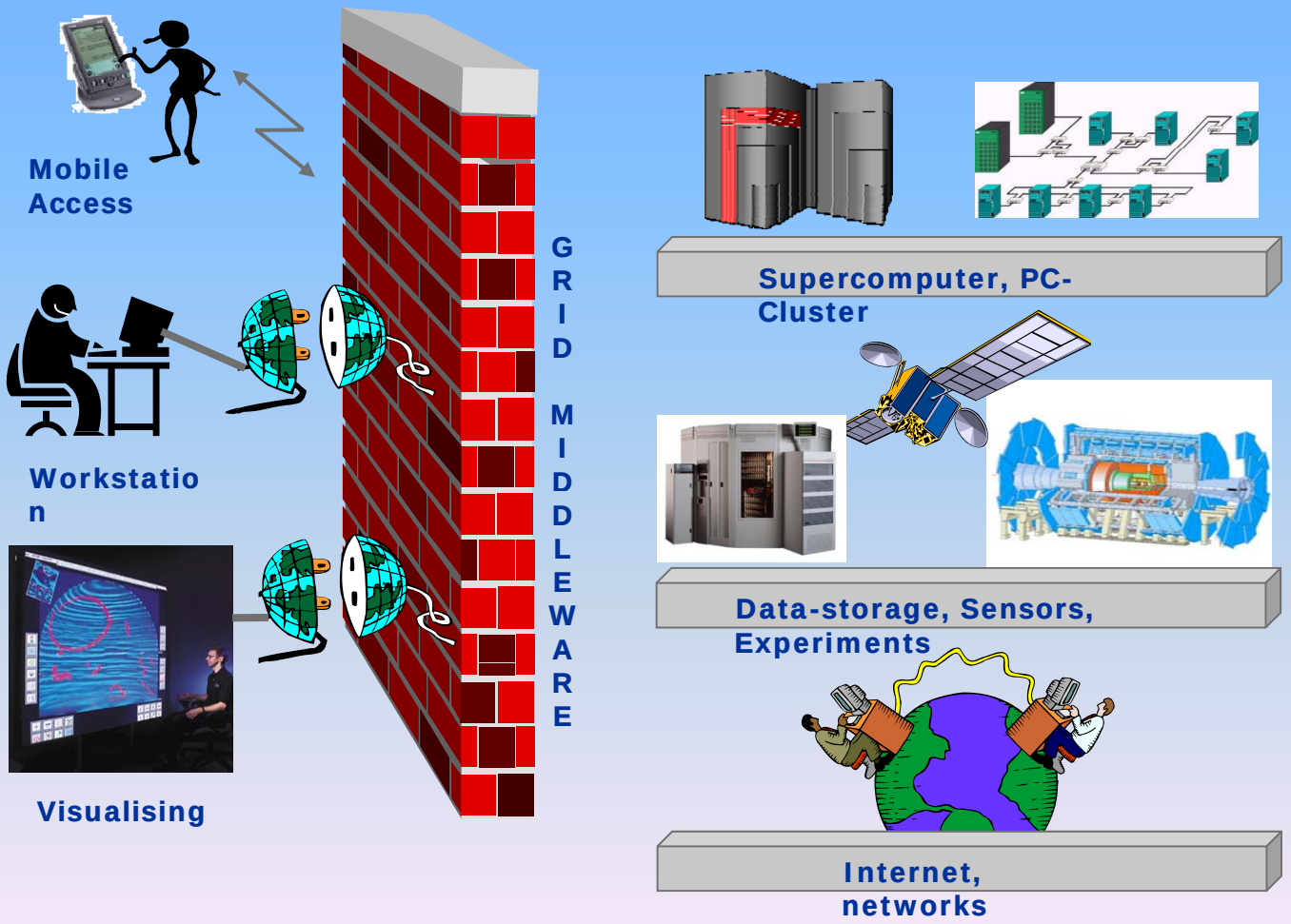
- Distributed resources
- No central control



EGEE: Tiers approach



Grid


Mobile Access
Workstation
Visualising
GRID MIDDLEWARE
Supercomputer, PC-Cluster
Data-storage, Sensors, Experiments
Internet, networks

EGEE infrastructure

Region	Countries	Sites	CPU
CERN	5	12	6400
UK/I	2	25	8384
Fr	2	12	7238
De/CH	2	15	4413
It	1	34	4341
NE	9	30	3289
SEE	8	38	2727
CE	7	24	2588
SWE	2	18	1938
A-P	8	20	1884
Ru	2	15	738
Totals	48	243	44040

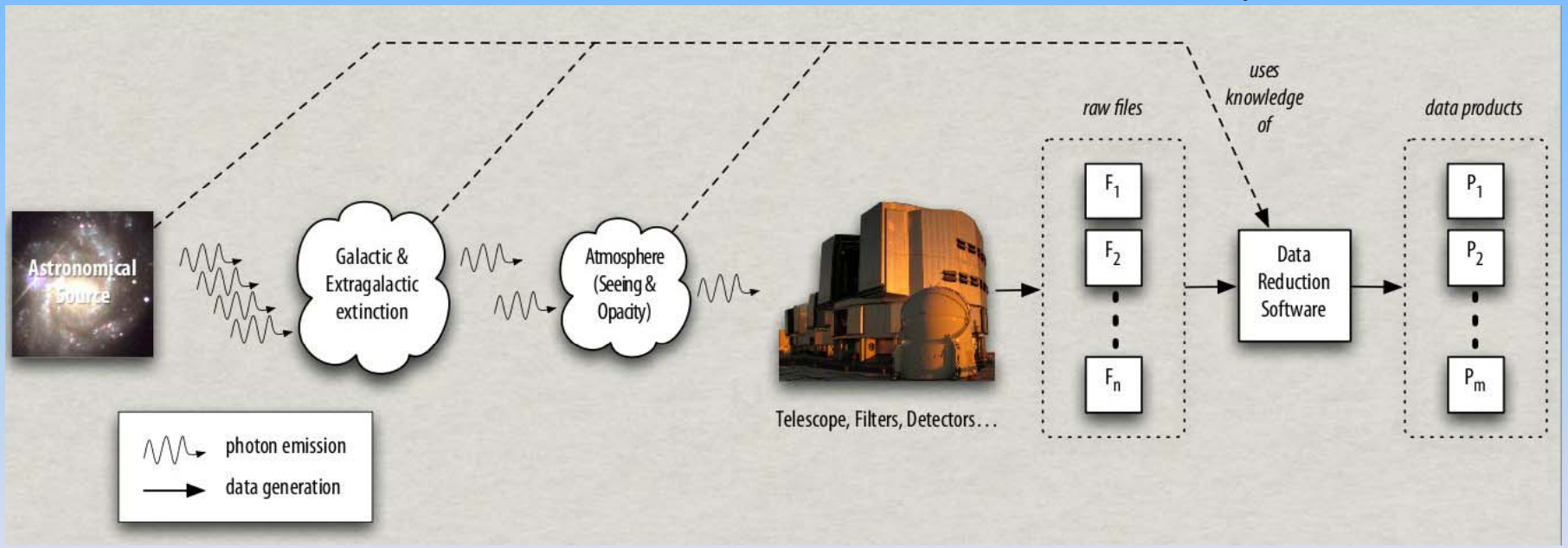
Grid in Netherlands: BigGrid

- 3 general sites
  - SARA
  - Nikhef
  - RuG
- 6 Life science sites
- About 3000 CPUs
- Several PB storage

VO: Data Provenance

J.Santander, A.Delgado, The case for Data Provenance
IVOA Interop 2009

http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/InterOpNov2009DM



Connect data reduction facilities (like Astro-WISE) with telescopes (raw data archive on telescope) and end-user (VO interfaces) by providing complete description of raw data (general data model for raw data).