The Astro-Wise Information system

From ASTRO WISE *to Wise technologies*

Data about data

Edwin A. Valentijn

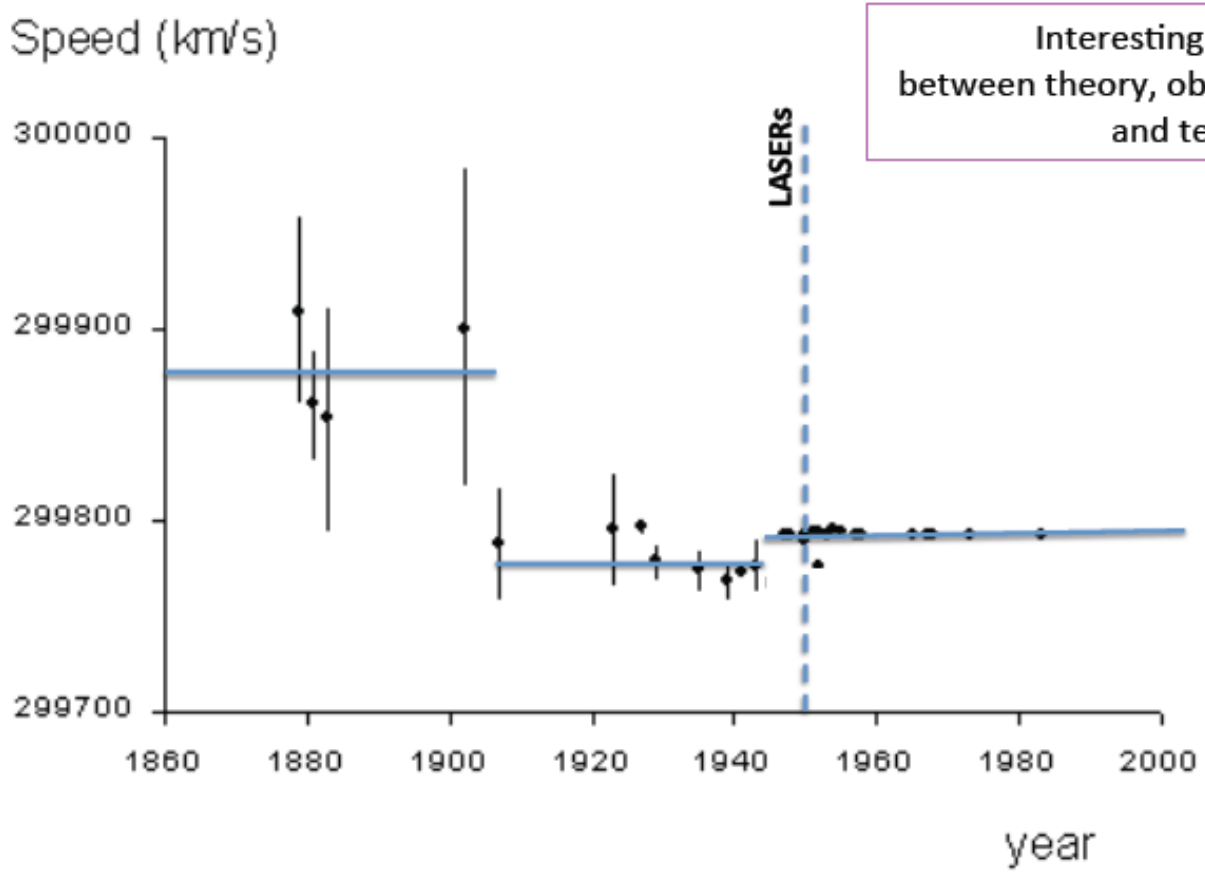
Prof Astronomical Information Technology

Target- OmegaCEN –Kapteyn

University of Groningen

SciOPS ESAC 12 Sept 2013

A Historical Digression: Speed of Light

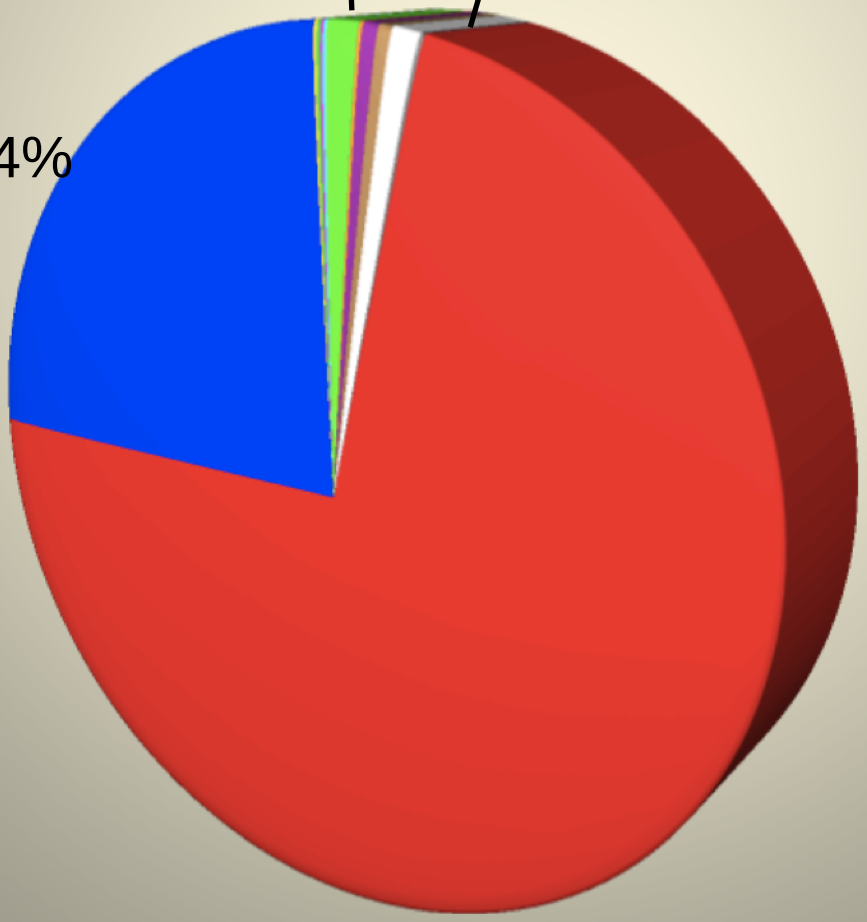


Interesting interplay between theory, observation and technology

The measurements we are to perform with Euclid are no less important or profound

Petabytes and beyond

FORS1 FORS2 GIRAFFE ISAAC MIDI
NACO SINFONI UVES VIMOS VISIR
VIRCAM OmegaCAM



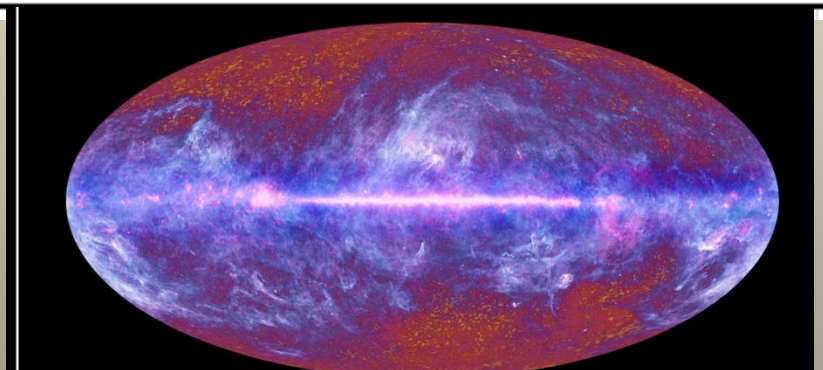
Peta -100 Peta

bits	#states		Byte
0	1e+00	1	pre Big Bang
1	2e+00	2	Big Bang
8	3e+02	256	Machu Pichu
16	7e+04	65536	
24	2e+07	16777216	Mega
32	4e+09	4294967296	Giga
40	1e+12	1099511627776	Tera
48	3e+14	281474976710656	
56	7e+16	72057594037927936	Peta
64	2e+19	18446744073709551616	100 Peta

Data about Big Data

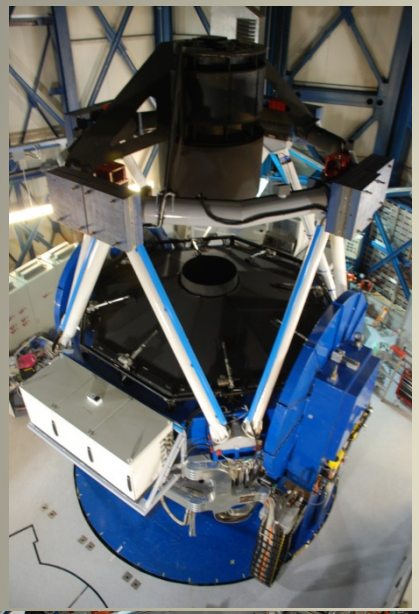
The Universe

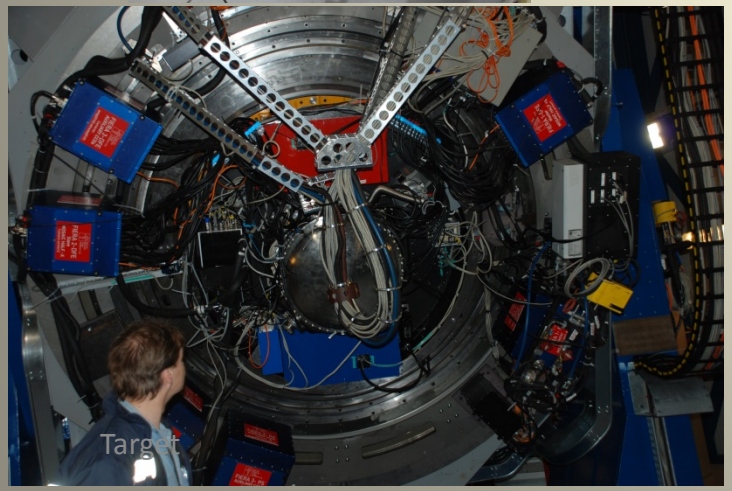
bits	#states		Byte
0	1e+00	1	pre Big Bang
1	2e+00	2	Big Bang
8	3e+02	256	Machu Pichu
16	7e+04	65536	
24	2e+07	16777216	Mega
32	4e+09	4294967296	Giga
40	1e+12	1099511627776	Tera
48	3e+14	281474976710656	
56	7e+16	72057594037927936	Peta
64	2e+19	18446744073709551616	100 Peta
128	3e+38	340282366920938463463374607431768211456	
256	1e+77	115792089237316195423570985008687907853269984665640564039457584007913129639936	



TarGet

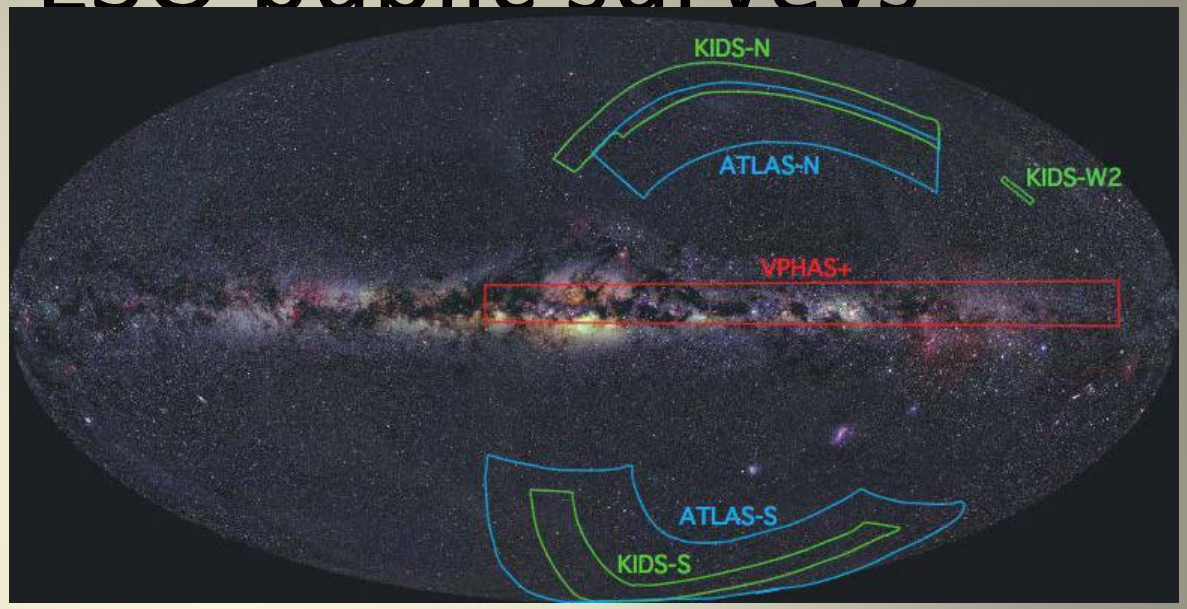
98 99 100 101 102 103 104

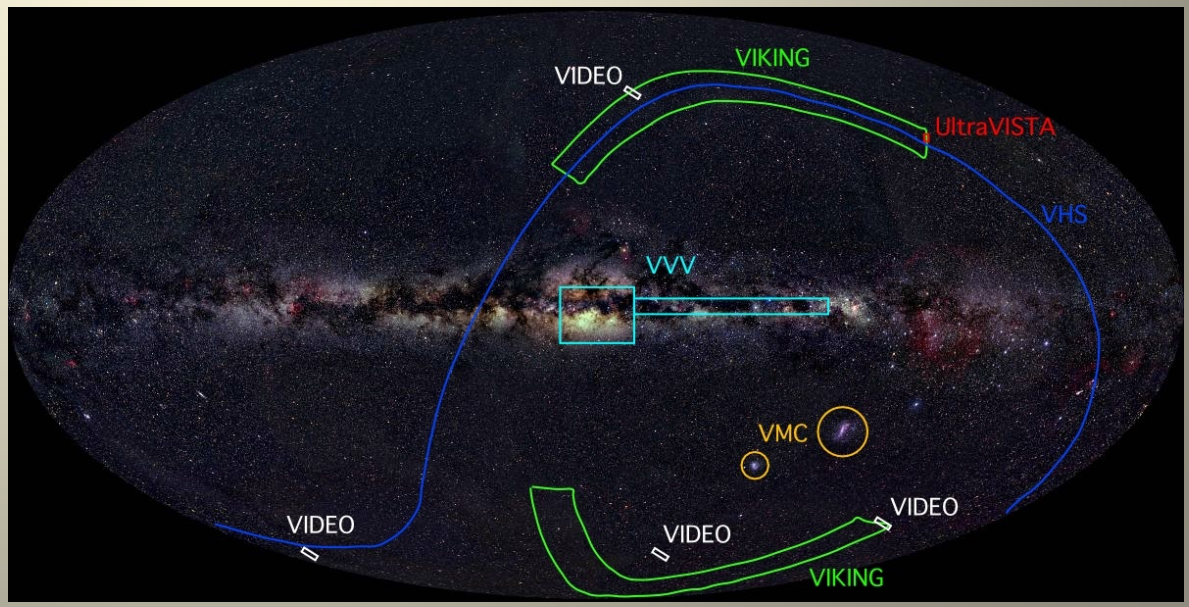


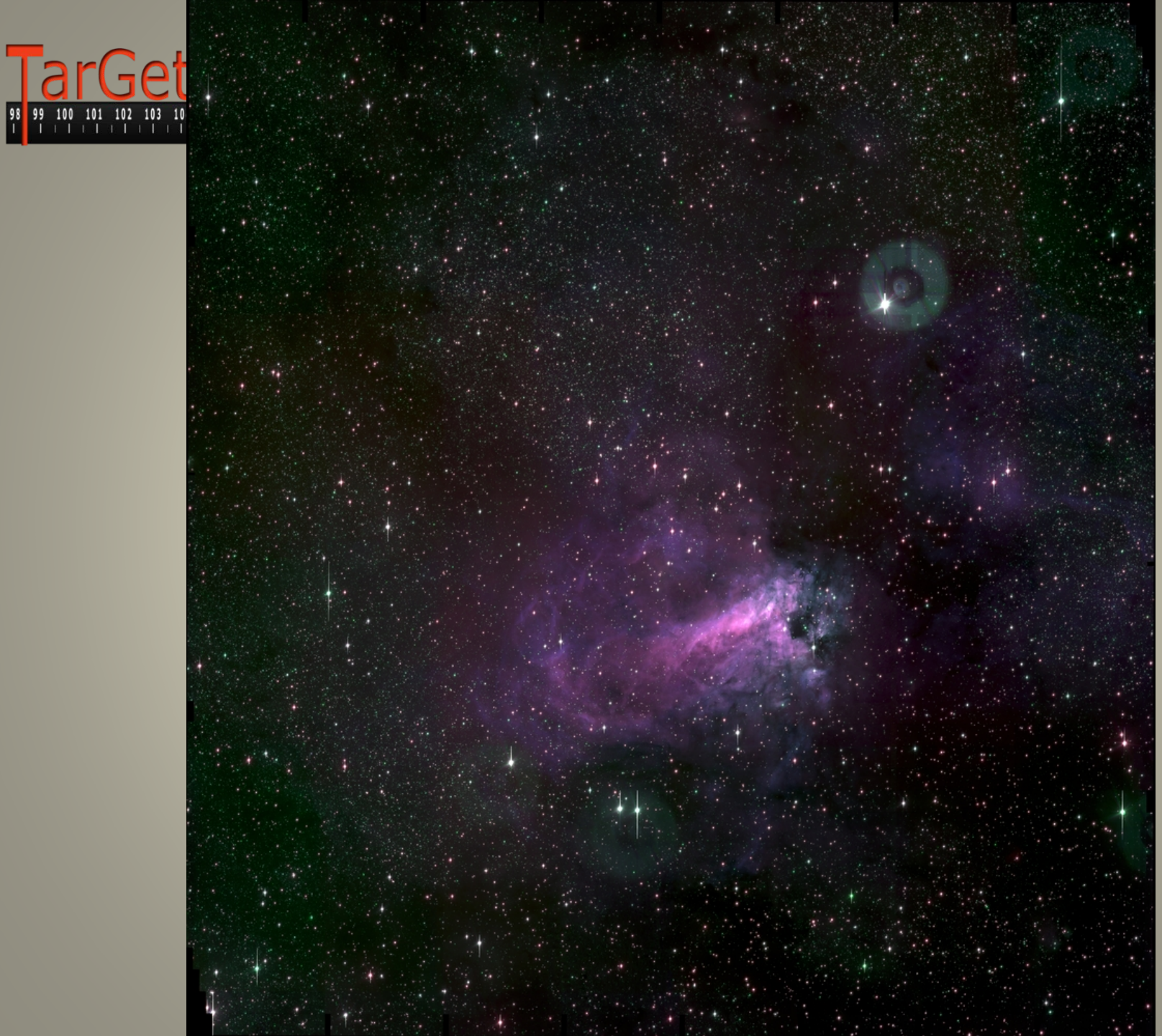


ESO public surveys

- VST:
- ATLAS
- VPHAS+
- KiDS
- VISTA:
- VHS
- VIKING
- VIDEO
- UltraVISTA
- VIDEO
- VVV





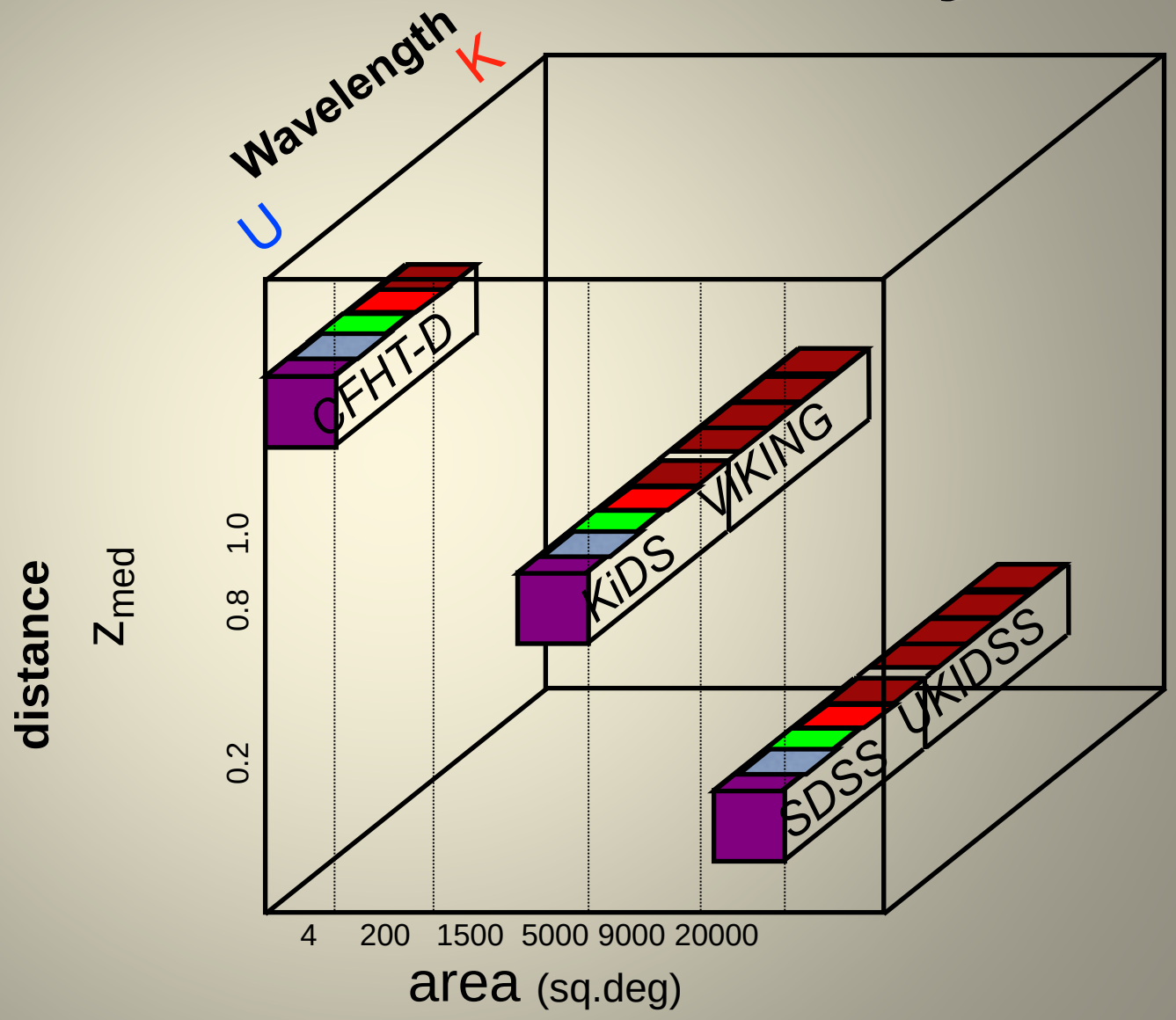

TarGet
98 99 100 101 102 103 10

KiDS vs. other surveys

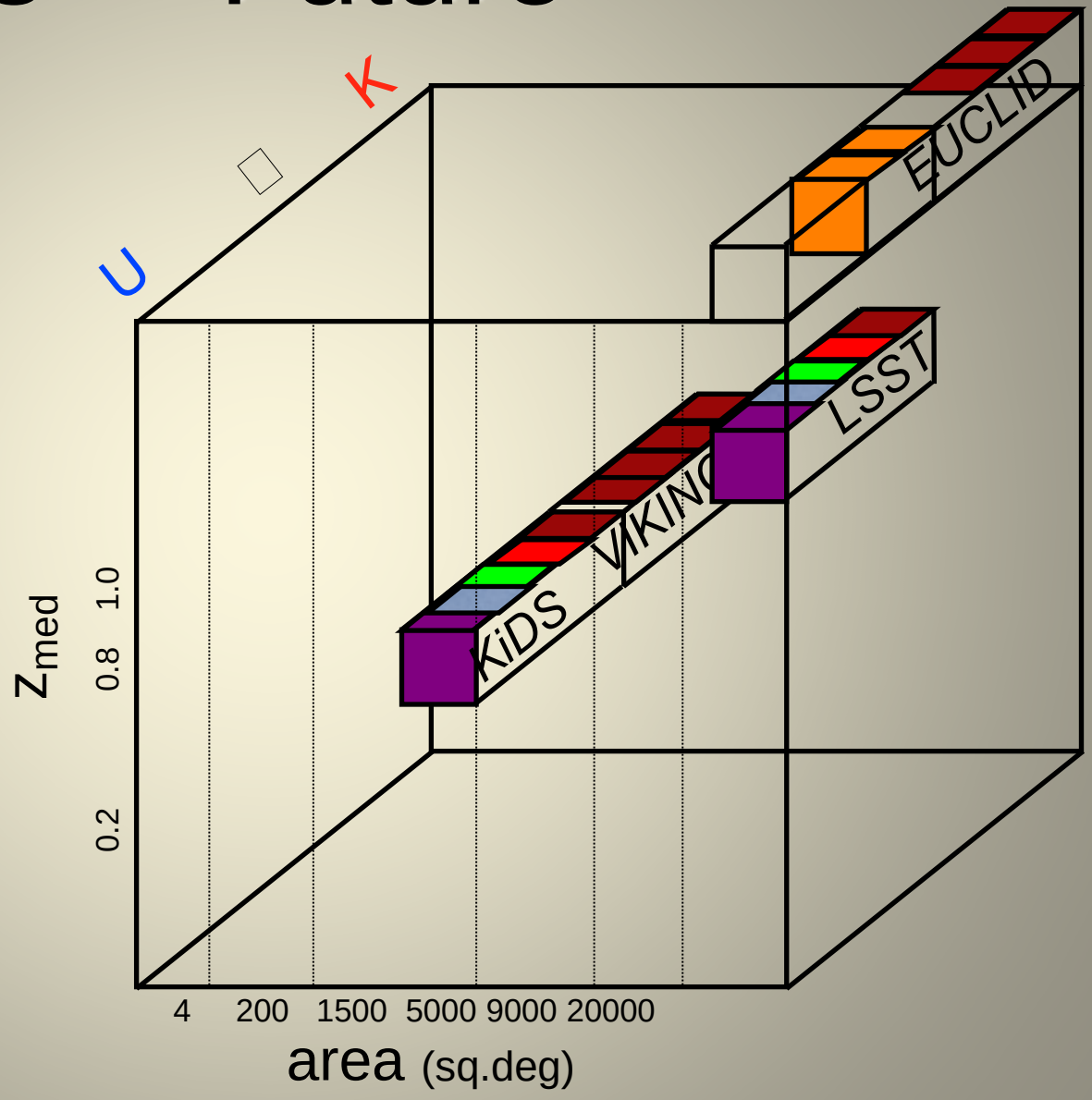
Earlier:

SDSS

CFHT-D



Kids - Future


u
K
EUCLID
LSST
VIKING
KiDS
z_{med}
0.2
0.8
1.0
4
200
1500
5000
9000
20000
area (sq.deg)

TarGet

98 99 100 101 102 103 104

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

try astro-wise

- 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

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 queries[1]: 514980167

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

Status of services at Astro-WISE nodes

Bonn
Groningen
Leiden Nijmegen
München
Napoli

poll details

Updated: 24 Sep 2011 01:42:02

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

AstroWise -> Wise

- 64 bit identifyer
- Terabytes of pointers
- Extreme datalineage ++
- Modeling dependencies
- Distributed

Experimental Astronomy
Volume 35 januari 2013
Topical issue on AstroWise
19 papers 389 pages

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

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)

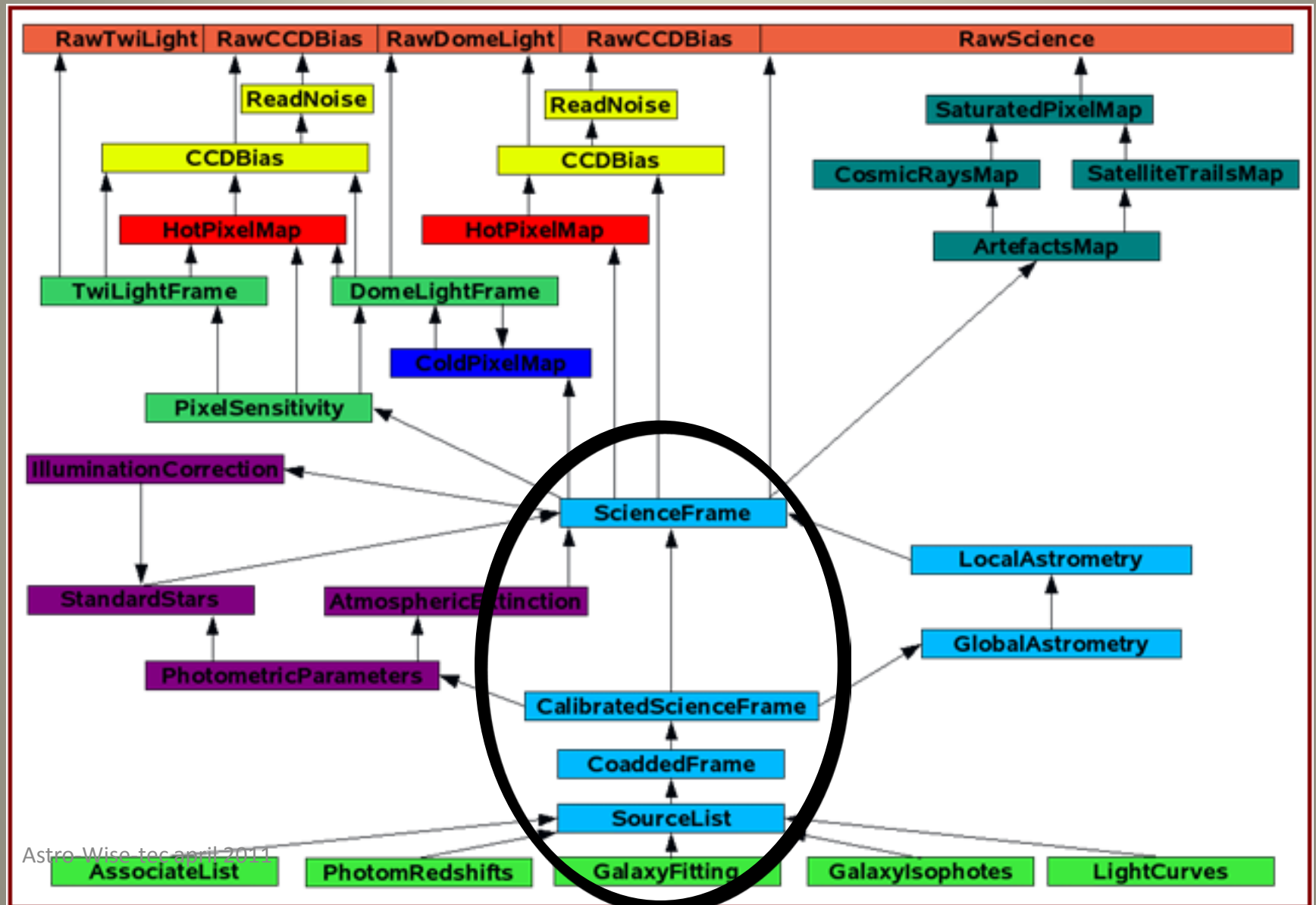
FileServer Server (Python)

RAC (OmegaCEN)

Oracle Server

© 2003 Astro-Wise

TARGET diagram



Target processing: ++ the make metaphor

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

The processing chain is

ReadNoise <-- Bias <-- HotPixels

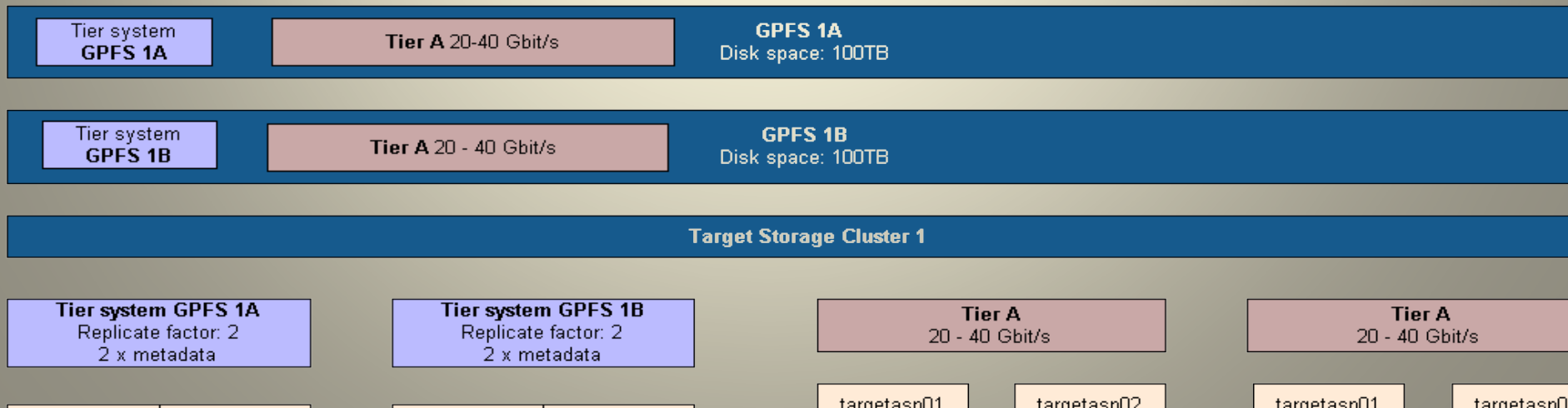
- > class HotPixelMap(ProcessTarget):
- > > def self.make()

- > 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

10 Peta byte Target testbed with IBM GPFS

- Tiered storage system – 4 levels
- Data about Big Data - GPFS
- 400 Million files – big and small
- Research collaboration -> Upgrades
  - Performance: hardware, software, design



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: awpverdoes project: WFI@2...

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

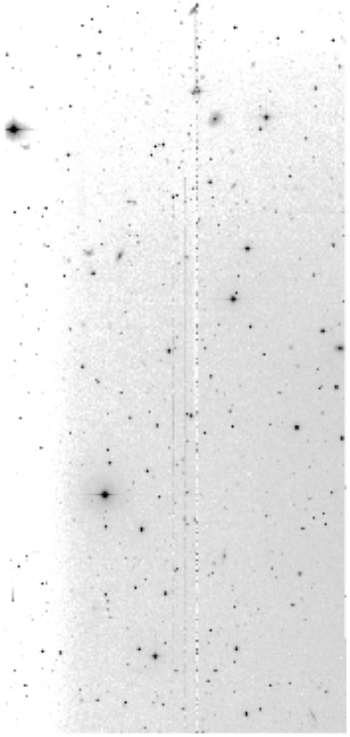
creation_date	2008-03-29 20:15:06
is_valid	1
quality_flags	0
zeropoint	24.759
zp_error	0.000
zp_origin	derived
num_sources	173

Observational Details

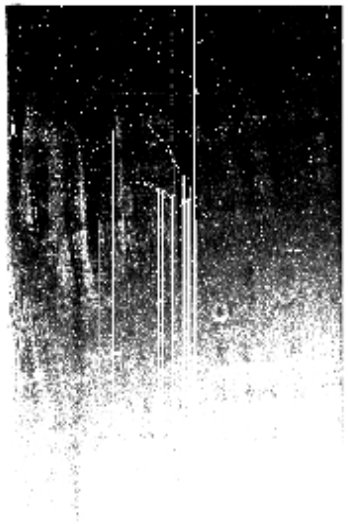
DATE_OBS	2002-03-18 03:35:21	OBSERVER	UNKNOWN
MJD_OBS	52351.1495509	EXPTIME	299.9176
OBJECT	She-W	AIRMSTRT	1.246
R.A.	13:25:30.0000	AIRMEND	1.246
Dec.	-31:35:44.6808	Filter	#842
		mag_id	JohnsonB

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

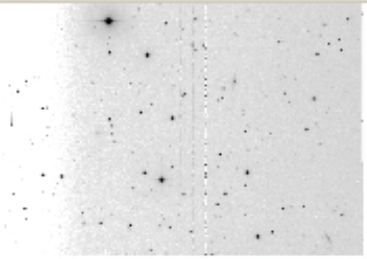
Target

Done

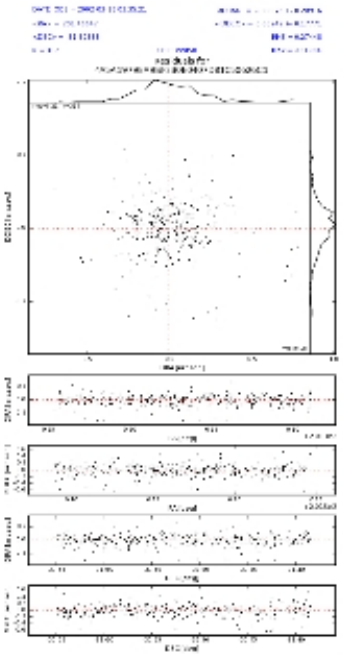
File Edit View History Bookmarks Tools Help
http://quality.astro-wise.org/QualityWISE?object_str=RegriddedFrame&object_id=4A5A1A6C90048502E0407D
Google

Photometry Details

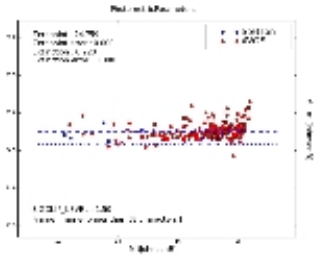
creation_date	2008-03-29 20:15:06
is_valid	1
quality_flags	0
zeropoint	24.759
zp_error	0.000
zp_origin	derived
num_sources	173
extinction	0.220
ext_error	0.000



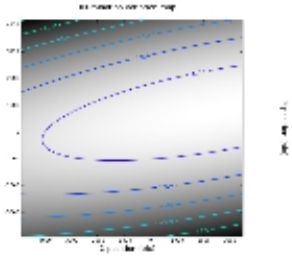
AstrometricParameters



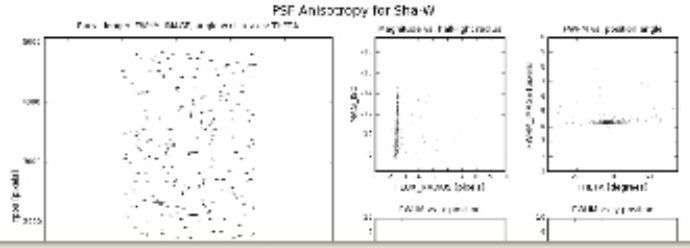
PhotometricParameters



IlluminationCorrection



PSF Anisotropy

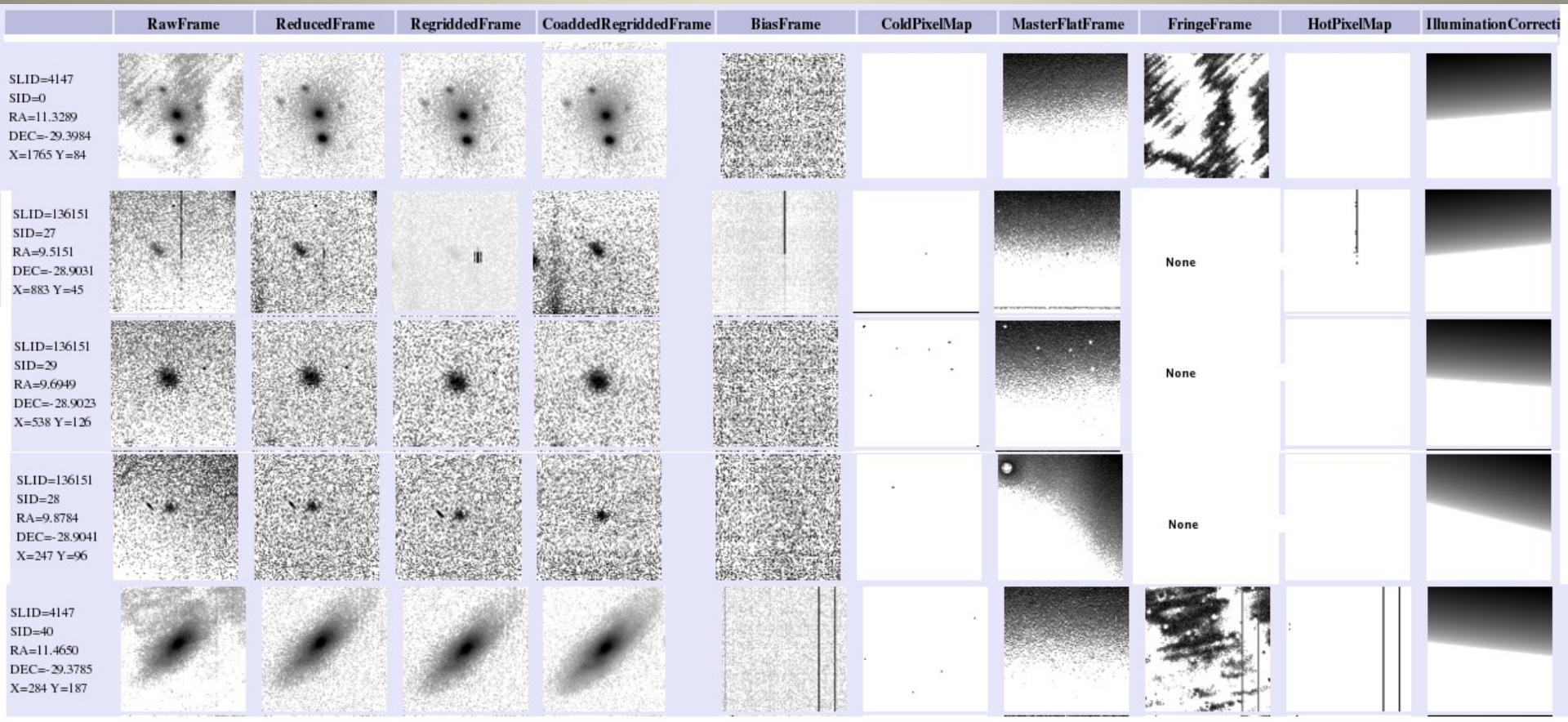


ASTRO WISE

Target

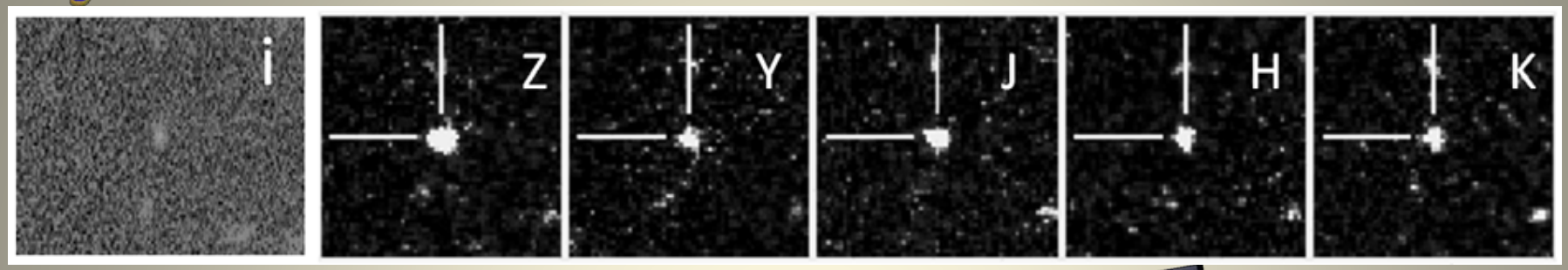
Done

Extreme data lineage

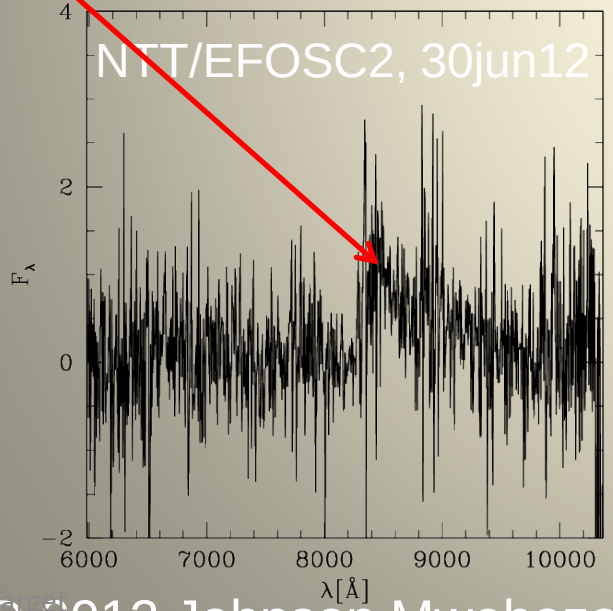


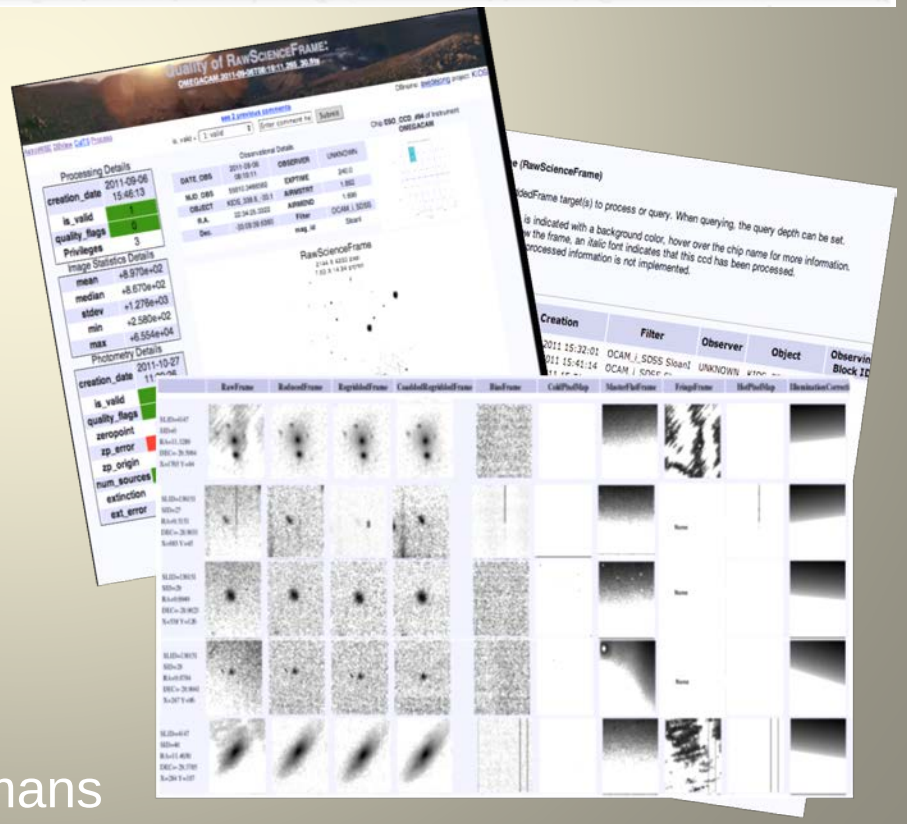
Most distant quasars in the universe

KiDS QSO at z~5.8 in 5 months



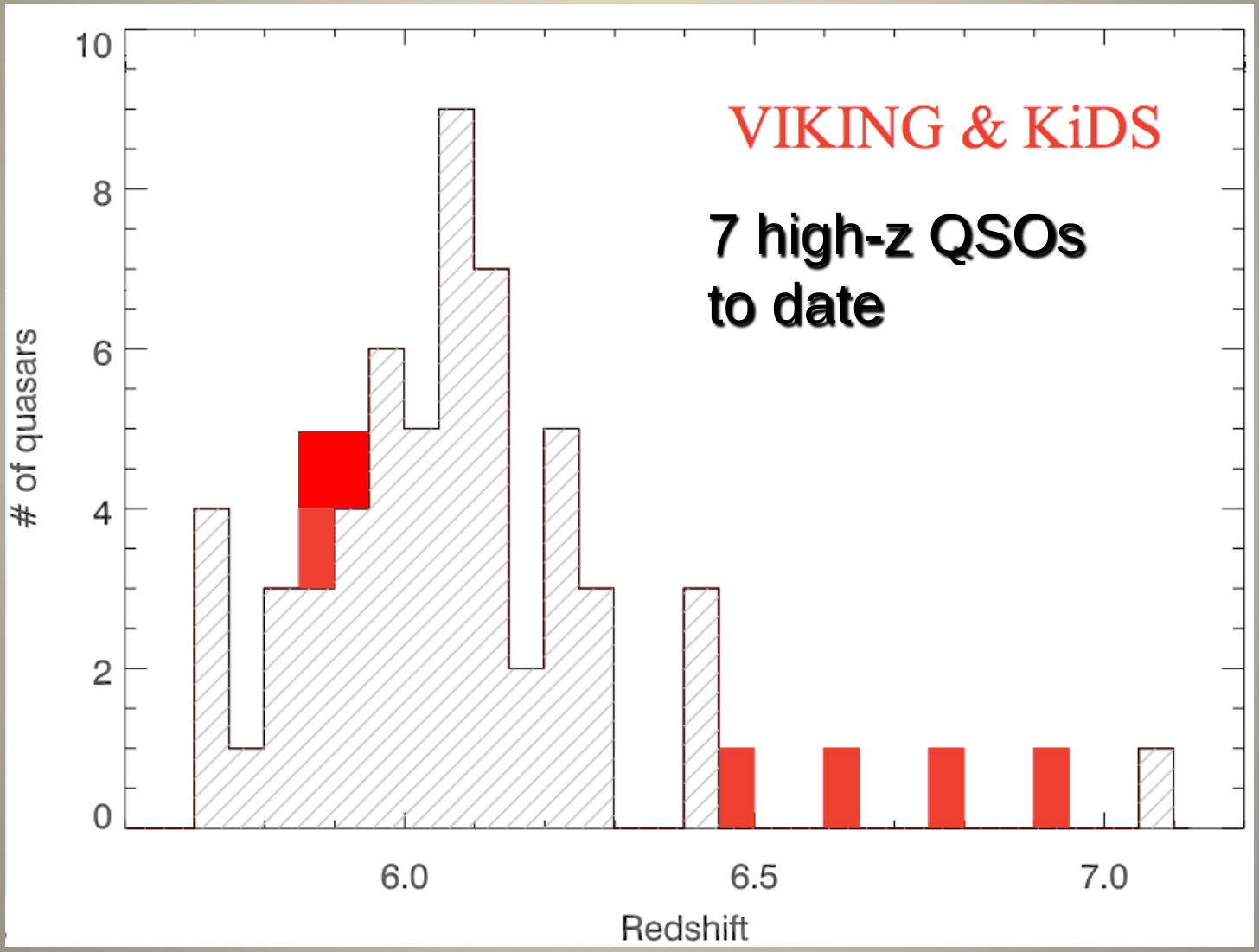
Ly-break+alpha





PhD 2012 Johnson Mwebaze / Venemans

KiDS: first results



Query-Driven Visualization

Bridging the gaps between Processing, Archiving and Analyzing

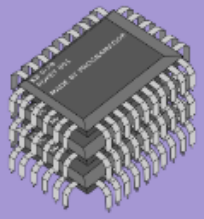
H. Buddelmeijer, buddel@astro.rug.nl
Kapteyn Astronomical Institute, University of Groningen

university of groningen
faculty of mathematics and natural sciences
kapteyn astronomical institute

Traditional 'Pushing' Approach

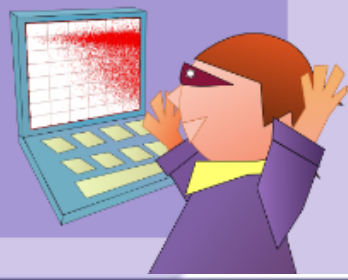
1 Raw Archive

2 Processing



3 Storage

4 Analysis



Query-Driven 'Pulling' Approach

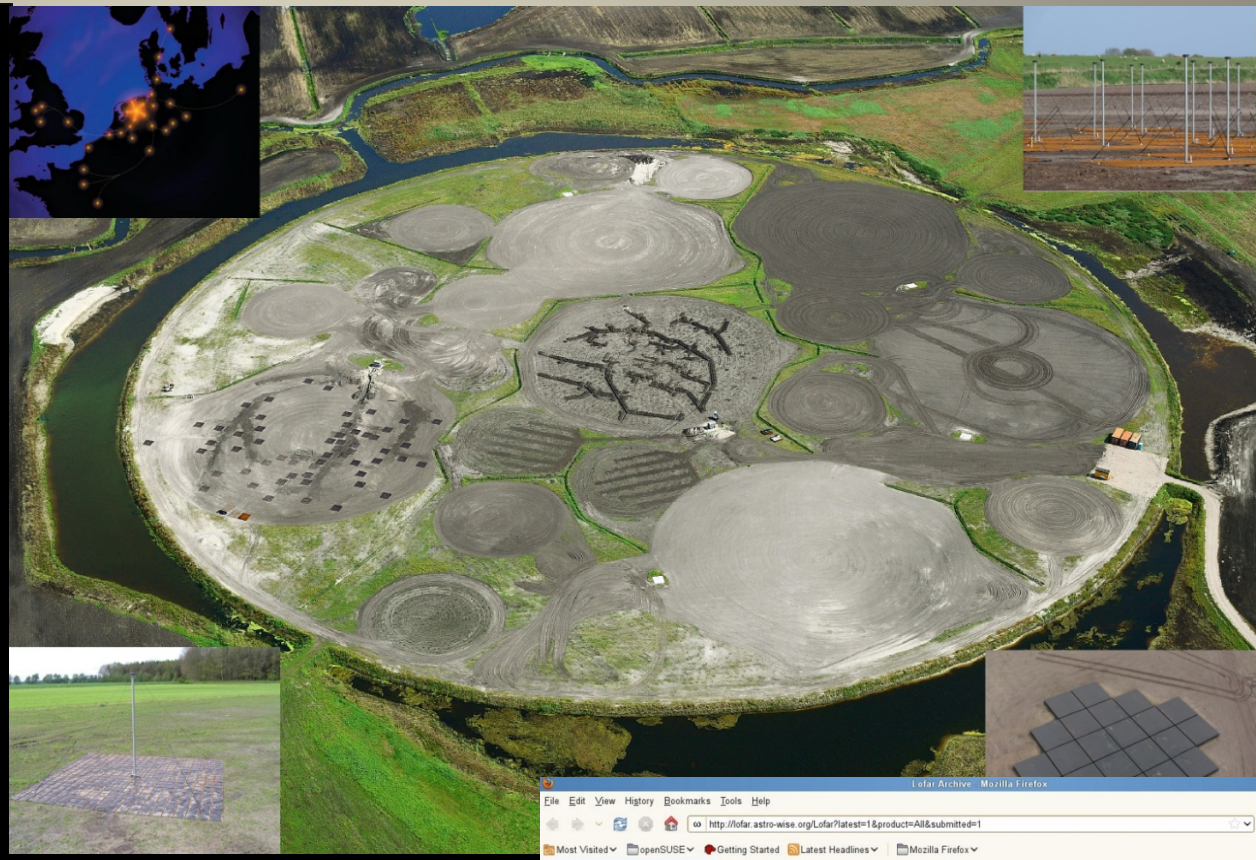
4 Raw Archive

3 Processing

2 Storage

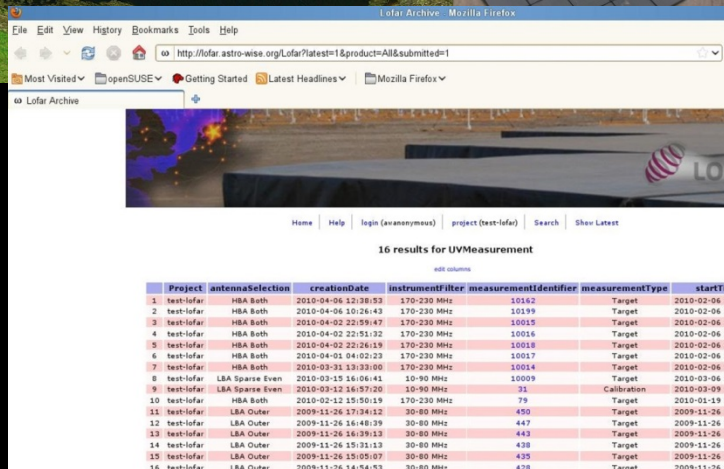
1 Analysis

Astro-WISE – LOFAR LTA



IBM- Blue Gene/P

5 Petabyte/year



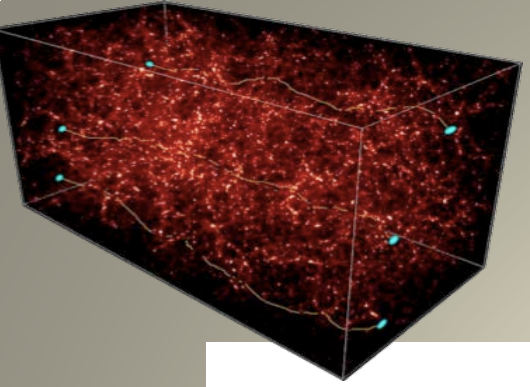
	Project	antennaSelection	creationDate	instrumentFilter	measurementIdentifier	measurementType	startT
1	test-lofar	HBA Both	2010-04-06 12:38:53	170-230 MHz	10162	Target	2010-02-06
2	test-lofar	HBA Both	2010-04-06 10:26:43	170-230 MHz	10199	Target	2010-02-06
3	test-lofar	HBA Both	2010-04-02 22:59:47	170-230 MHz	10015	Target	2010-02-06
4	test-lofar	HBA Both	2010-04-02 22:51:32	170-230 MHz	10016	Target	2010-02-06
5	test-lofar	HBA Both	2010-04-02 22:26:19	170-230 MHz	10018	Target	2010-02-06
6	test-lofar	HBA Both	2010-04-01 04:02:23	170-230 MHz	10017	Target	2010-02-06
7	test-lofar	HBA Both	2010-03-31 13:33:00	170-230 MHz	10014	Target	2010-02-06
8	test-lofar	LBA Sparse Even	2010-03-15 16:06:41	10-90 MHz	10009	Target	2010-03-06
9	test-lofar	LBA Sparse Even	2010-03-12 16:57:20	10-90 MHz	31	Calibration	2010-03-09
10	test-lofar	HBA Both	2010-02-12 15:50:19	170-230 MHz	79	Target	2010-01-19
11	test-lofar	LBA Outer	2009-11-26 17:34:12	30-80 MHz	450	Target	2009-11-26
12	test-lofar	LBA Outer	2009-11-26 16:48:39	30-80 MHz	447	Target	2009-11-26
13	test-lofar	LBA Outer	2009-11-26 16:39:13	30-80 MHz	443	Target	2009-11-26
14	test-lofar	LBA Outer	2009-11-26 15:31:13	30-80 MHz	438	Target	2009-11-26
15	test-lofar	LBA Outer	2009-11-26 15:05:07	30-80 MHz	435	Target	2009-11-26
16	test-lofar	LBA Outer	2009-11-26 14:54:53	30-80 MHz	428	Target	2009-11-26

MuseWise
Jarle Brinchmann

ESO VLT

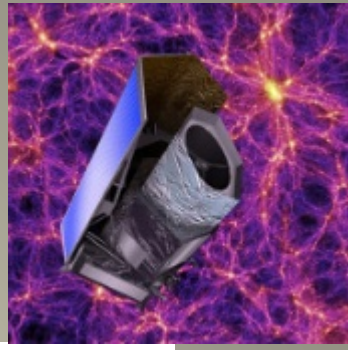
IFS 24 IFUs

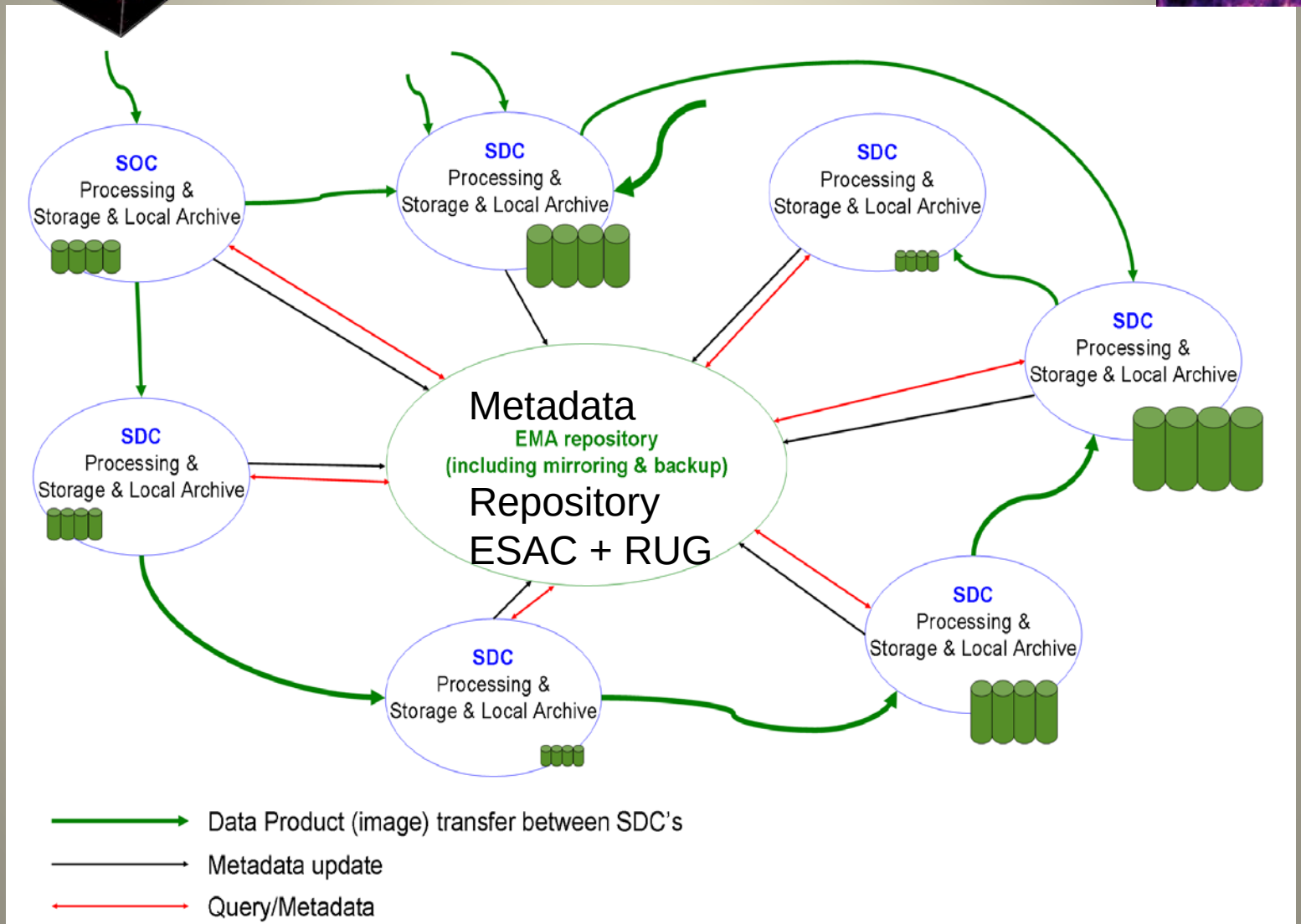
100 Gb/night



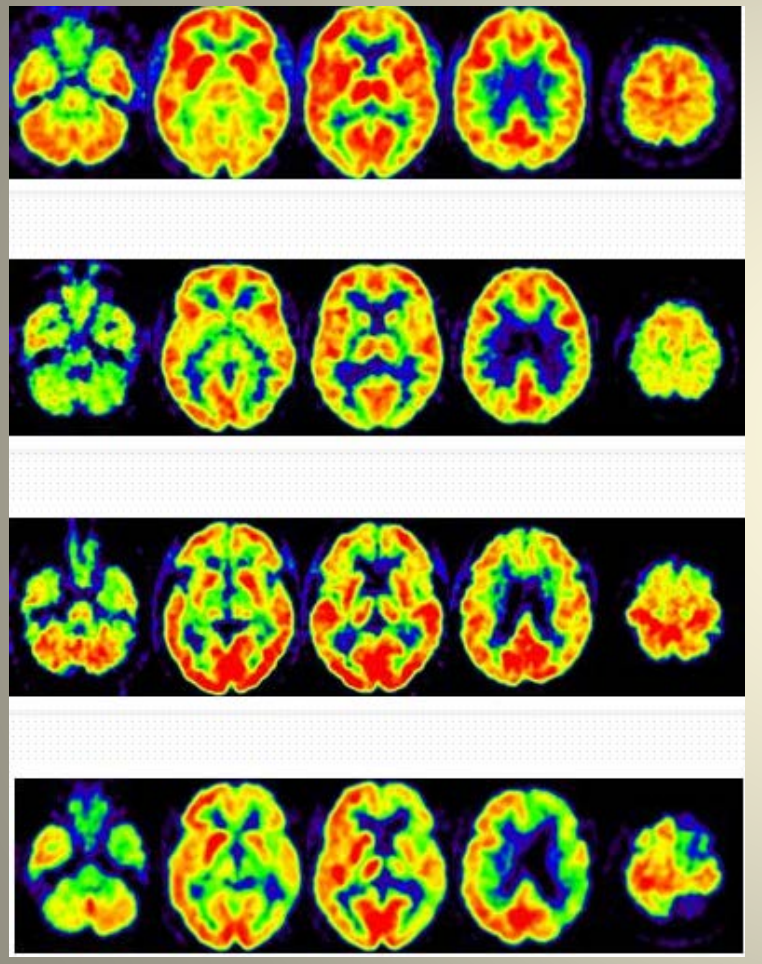
Euclid Archive System data centric approach

vs Planck – see Anna Gregorio





Wise



summary

Big Data

- Data centric approach
- Data about data
  - Extreme data lineage -64 bit pointers for everything
    - distribution
    - 500 Million file systems
    - Terabytes of pointers in databases

- Further R&D on
  - (distributed) visualisation
  - Visual analystics