Astronomical GRID Applications at ESAC

Christophe Arviset
Head of the Science Archives and Computer Engineering Unit

ASPERA Workshop
Lyon, 7th of October 2010

Presentation Overview

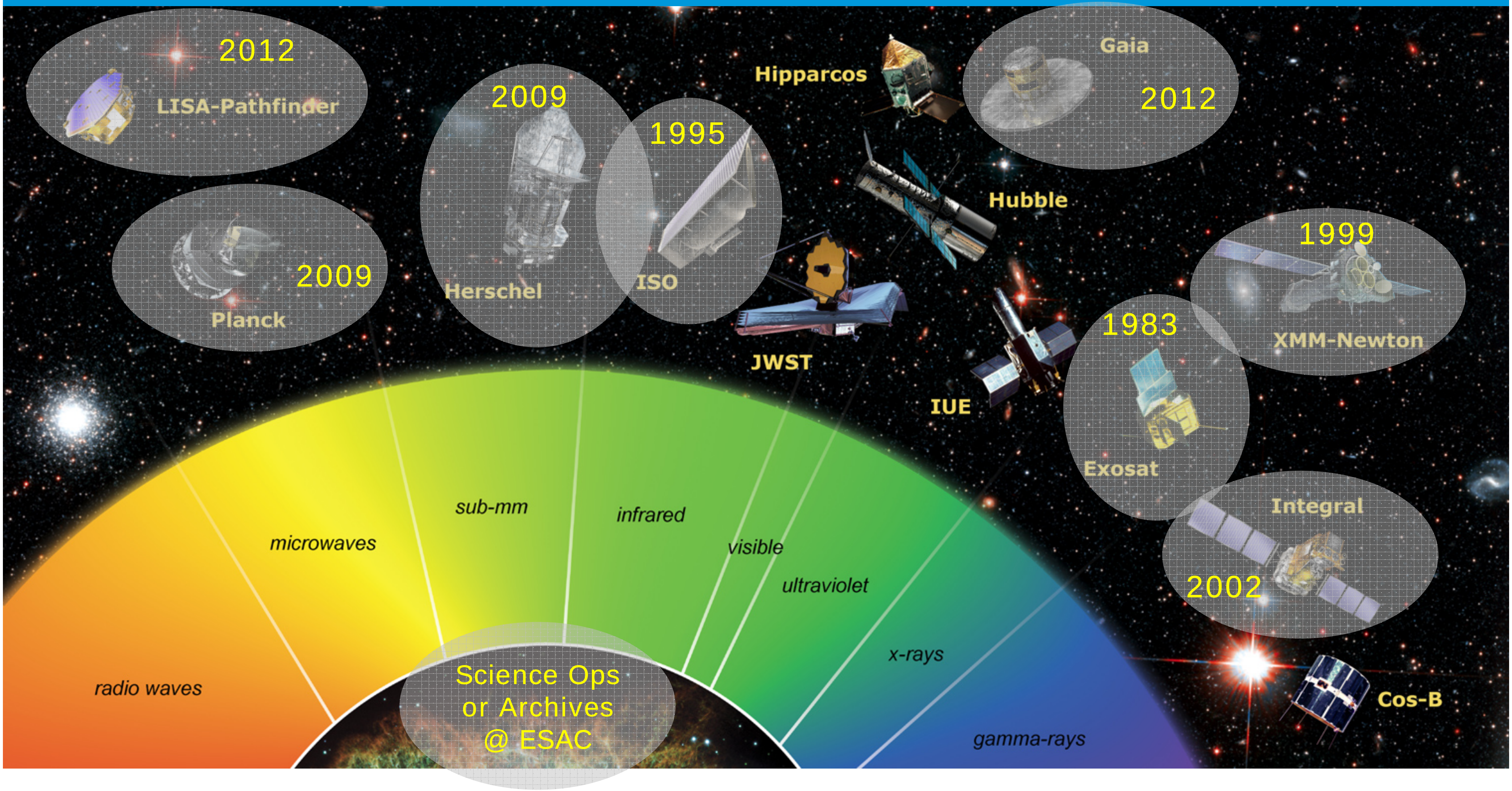
- ESAC and its GRIDs infrastructure
- Herschel Pipeline Data Processing
- XMM-Newton Remote Interface to Science Analysis
- XMM-Newton Calibration Monitoring
- Measure transverse motion of 730.000 stars
- XMM-Newton Mosaic creation
- Cloud and GRID : Future Directions

European Space Astronomy Centre

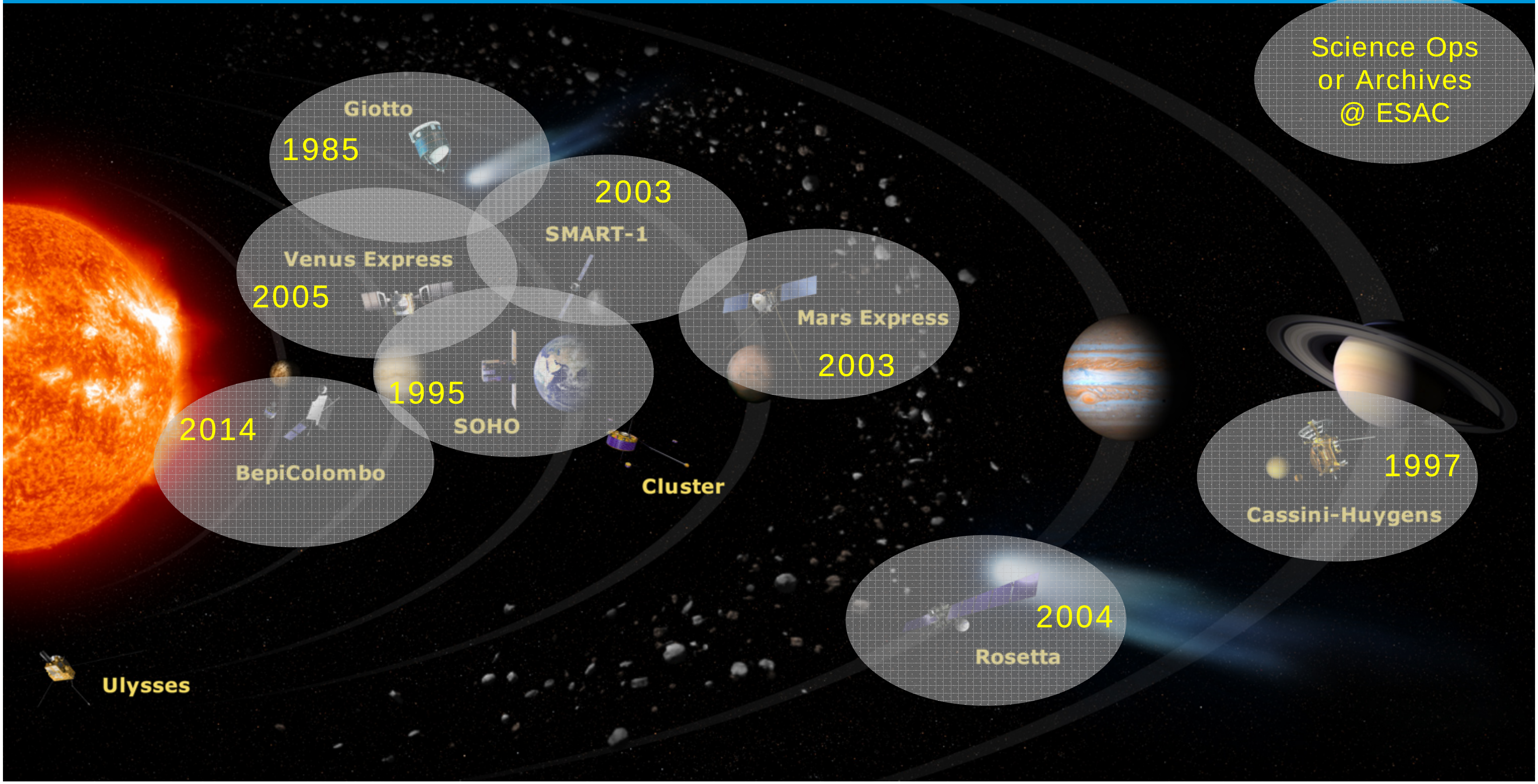
- ESAC default location for:
- Science operations,
  - long history with astronomical missions
  - Now also solar system missions
- Science archives,
  - Astronomy
  - Planetary and Solar Systems
- ESA Virtual Observatory activities,
  - ESAC is the European VO node for space-based astronomy.

http://www.esa.int/SPECIALS/ESAC/
Located near Madrid, Spain

The wonders of the Universe
Astronomy and fundamental physics mission at ESAC



Solar Systems Archives at ESAC



ESAC GRID Objectives

- ESAC to be a *Data Centre* and a *Data Processing Centre*
  - Natural link between Science Archives, GRID and VObs
- ESAC Scientists to use the GRID for their daily science tasks
  - Must be easy to use (“my homedir on the GRID”, “gridlogin”)
  - User support to port their software to the GRID
  - New massive computing facilities to substitute user workstation
- GRID in support to the projects
  - Project Standard Data Processing
  - On-the-fly reprocessing from the Archives
  - Instruments calibration monitoring and trend analysis
  - HPC in the GRID
- Collaboration with other GRID partners / organizations
  - EGI, Planck, UCM, IFCA, IVOA

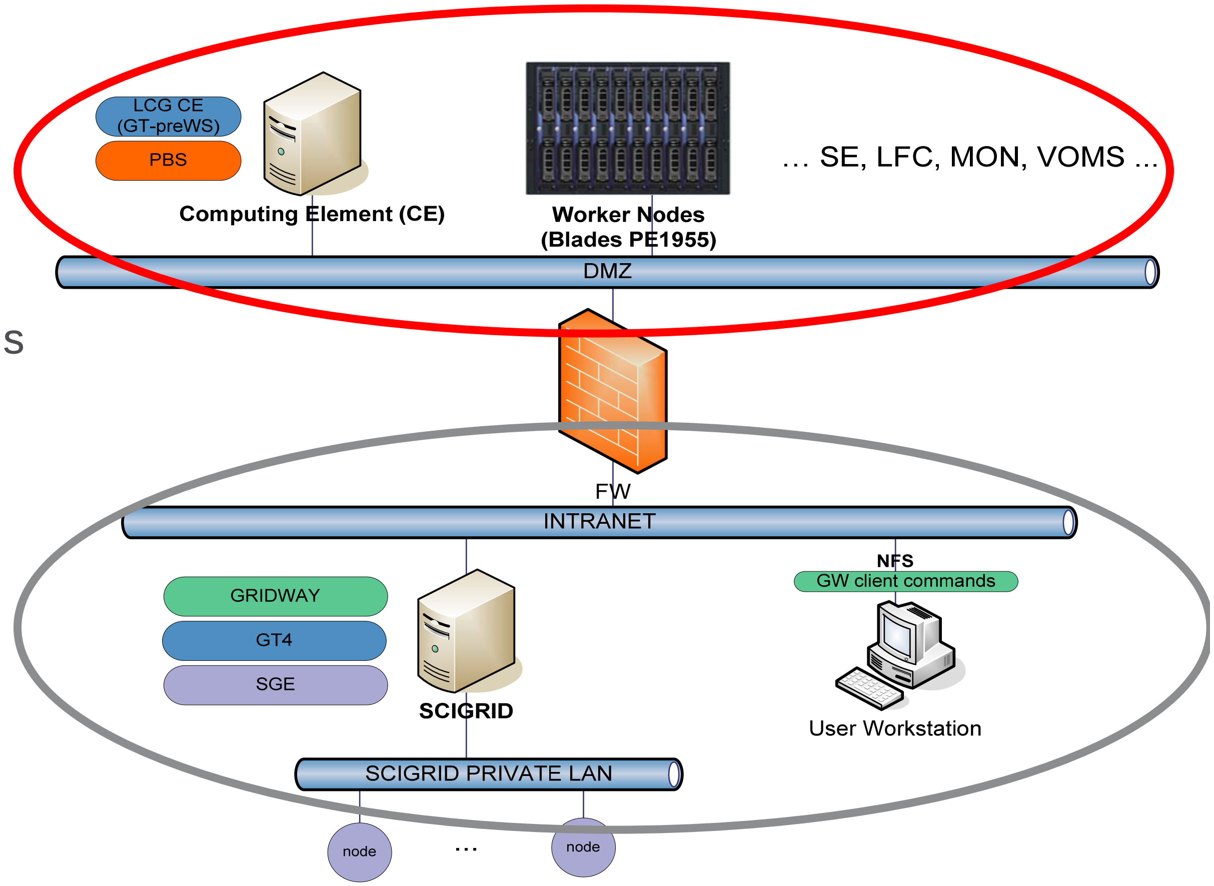
GRIDs at ESAC

ESACGRID

- On ESAC DMZ
- Part of EGEE, gLite based
- Close link to the ESAC Archives
- Project Data Processing
- Collaboration with other EGEE sites

SCIGRID

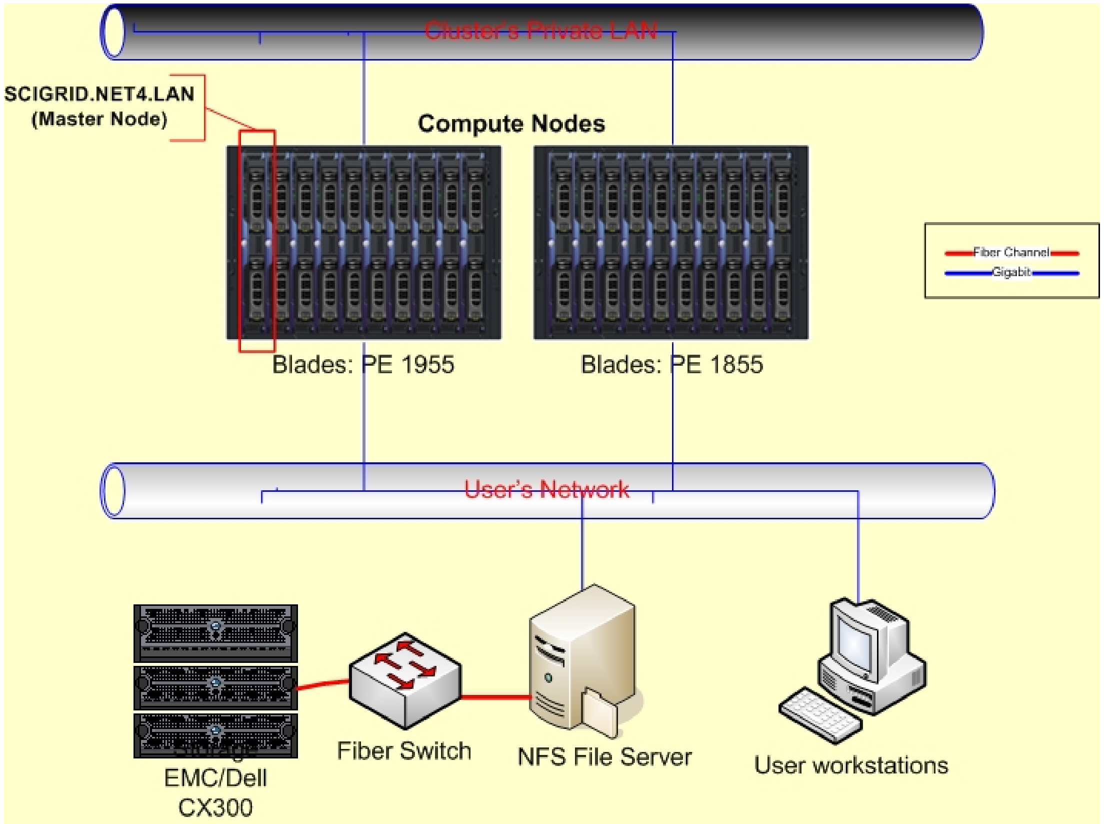
- On ESAC Intranet
- Globus based
- Scientist daily use
- Calibration monitoring



SCIGRID

SCIGRID

- On ESAC Intranet, fully integrated in users' network environment
- Cluster nodes mount through NFS
  - Daily science software
  - User's home directories
- Transparent access
  - Launch tasks to the Grid
  - Interactive session to the Grid ("glogin")
- In production since 2004



- 36 blades
- ~250 cores
- ~700 GB RAM
- RedHat

ESACGRID

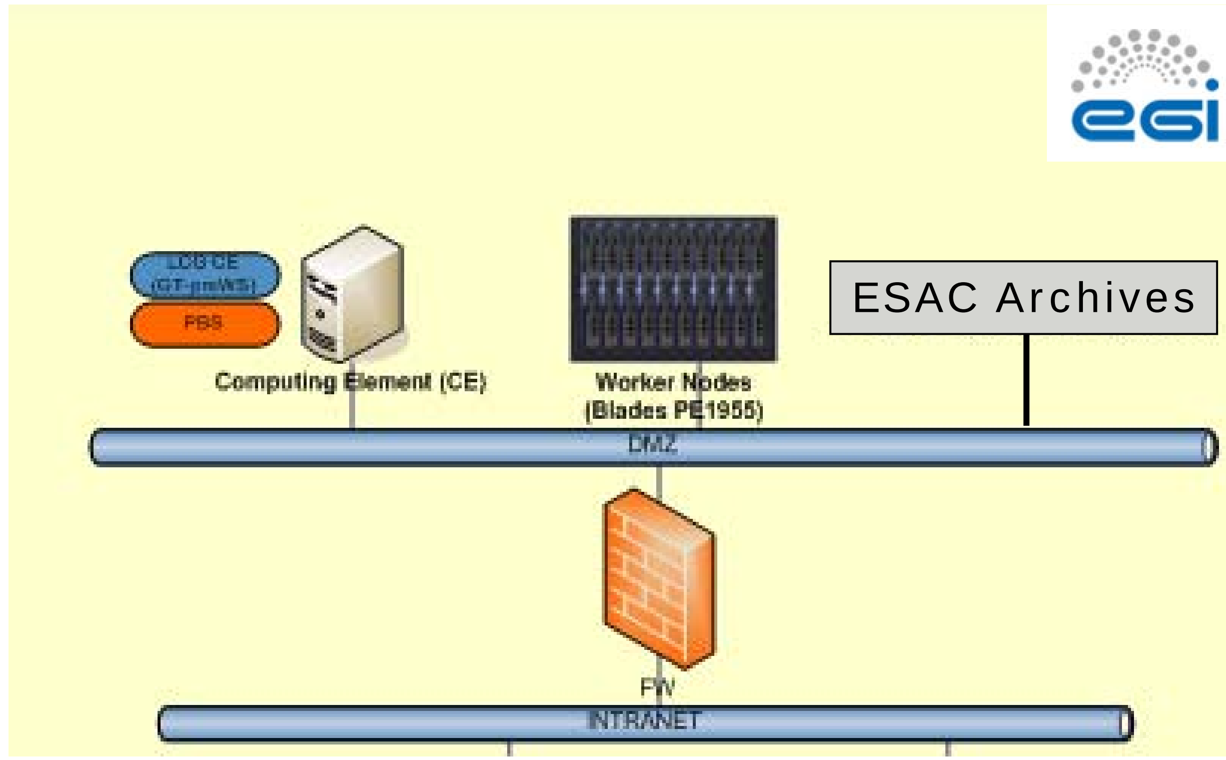
ESACGRID

- On ESAC DMZ
- Part of EGI, gLite based
- SGE job scheduler
- Scientific Linux

- Close to ESAC Science Archives
  - Faster access to the science data

- 56 blades
- ~400 cores
- ~1372 GB RAM

- In production since 2007

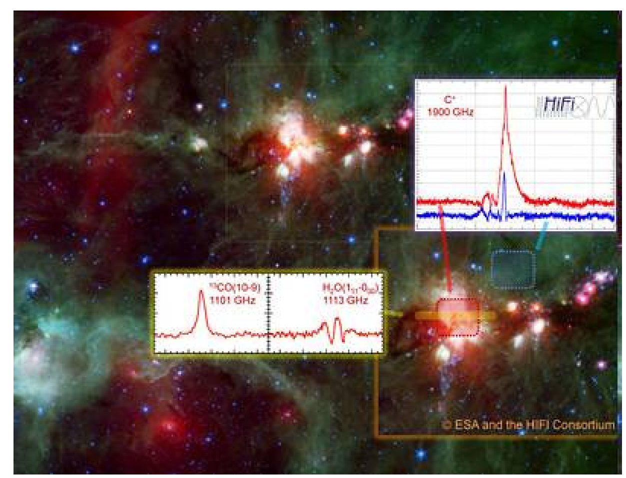


Herschel

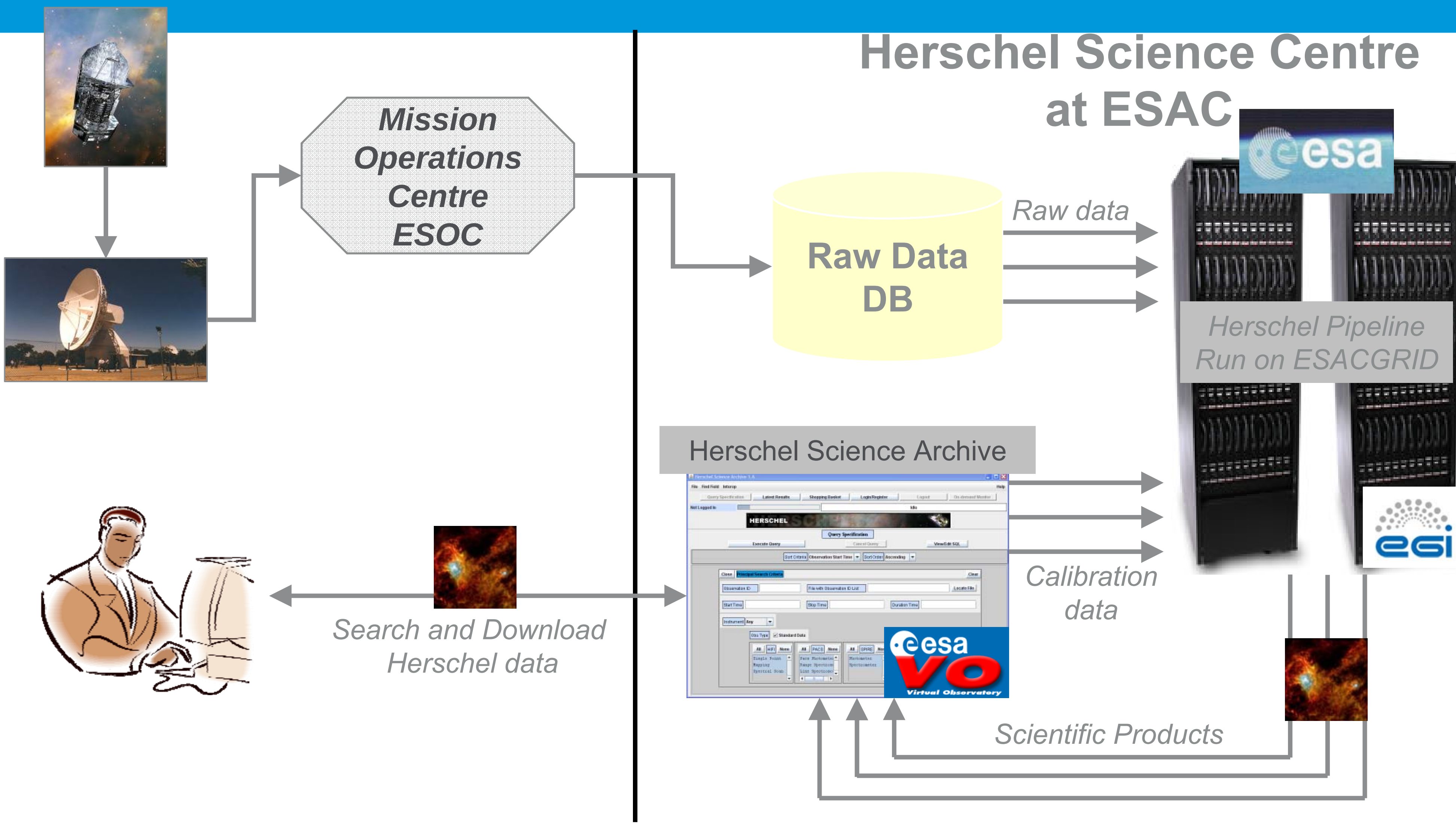
- Herschel is ESA's far Infrared observatory, launched in 2009

- Herschel Science Operations Centre @ ESAC is responsible of
  - Herschel Data Processing
    - Development with the instrument teams
    - Processing all Herschel raw data into science products
  - Herschel Science Archive (HSA)
  - (many other things)

- Herschel Data processing (raw data into science) requires
  - Systematic daily processing
  - On-the-fly reprocessing from the HSA
  - Regular bulk reprocessing of all raw data since launch
  - Goal : being able to process up to 20 days of data in 24h



Herschel Pipeline Data Processing



Advantages of Herschel DP on the Grid

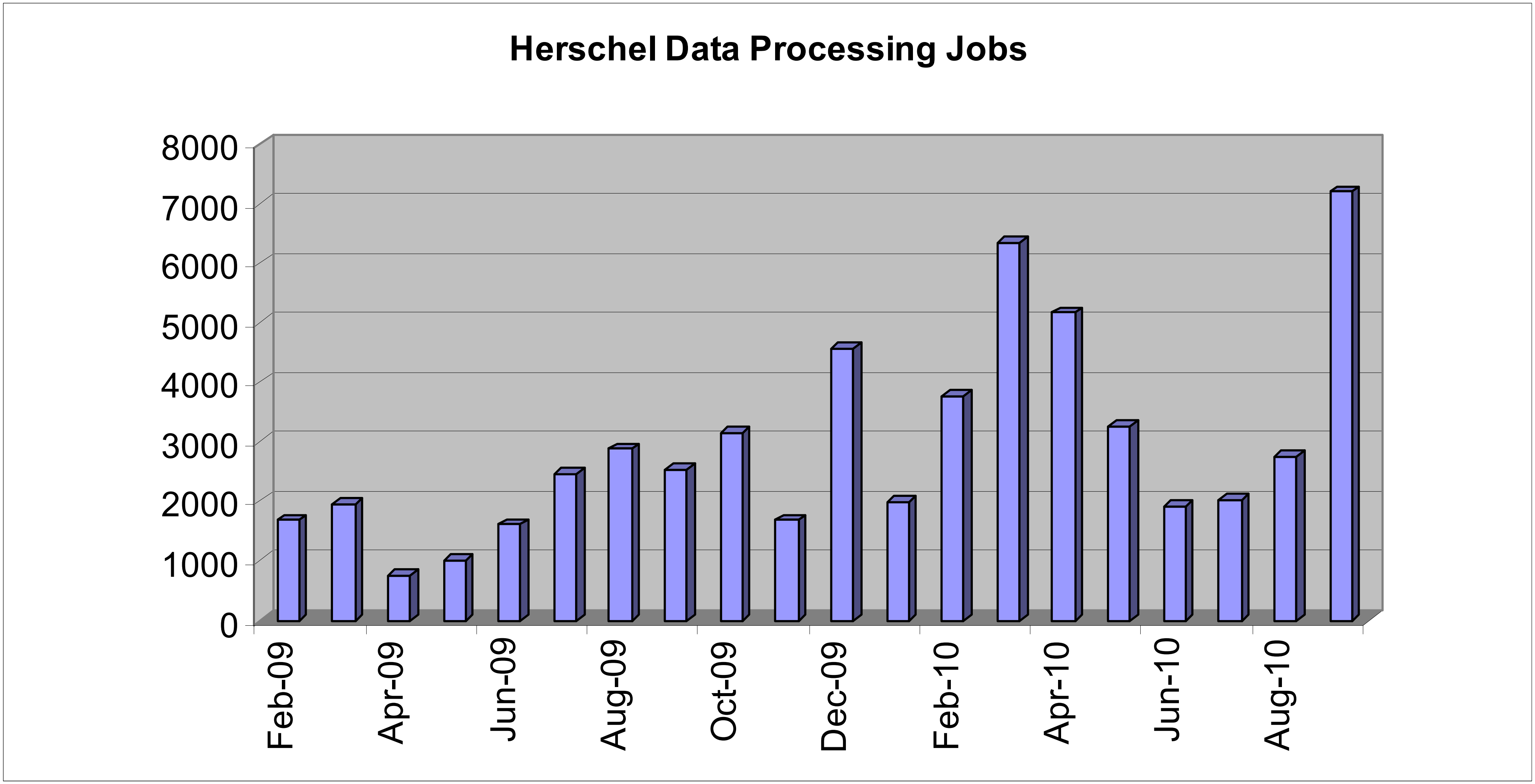
- Parallelization of tasks
  - Each observation can be sent on a GRID node
  - Many observations, days can be sent in parallel

- Robustness
  - HW/SW failures are automatically handled (usually simply restarting the job on a different node)

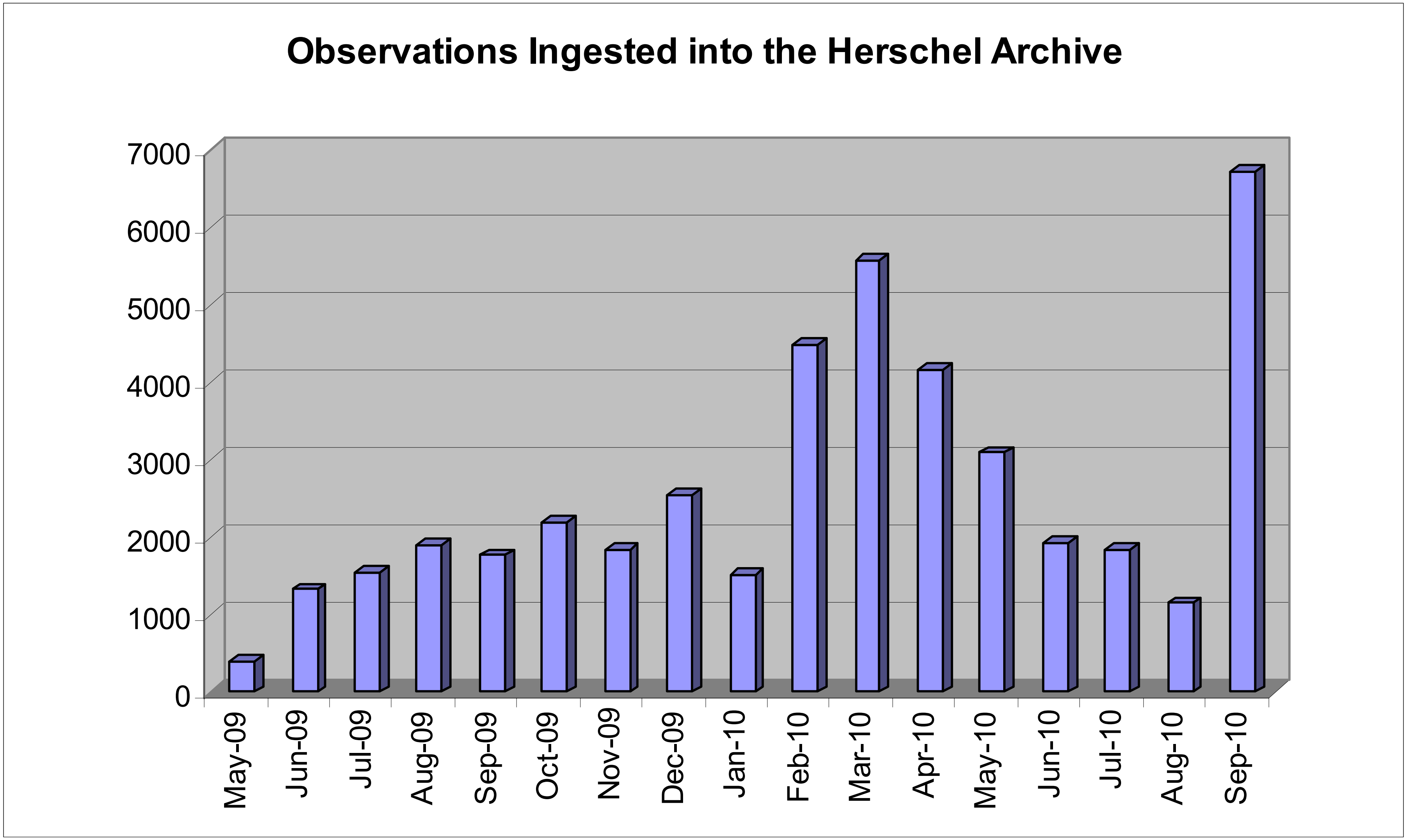
- Prioritization and resource management
  - Allow different queues with different nodes configuration (ie big RAM)

- Scalability
  - Grid can cope with standard processing, on-the-fly and bulk reprocessing
  - If more resources are required, we can add new Grid nodes

Herschel DP statistics - 1



Herschel DP statistics - 2



XMM-Newton Science Analysis Software (SAS)

- XMM-Newton is ESA's X-ray observatory since 1999
- XMM-Newton Science Operations Centre @ ESAC, together with Leicester Survey Science Centre maintain XMM-Newton Science Analysis Software (SAS), collection of tasks, scripts and libraries, specifically designed to reduce and analyze data
- SAS Freely distributed suite of programs ("tasks")
  - ~2 M lines of code, C++, F95, Perl
  - Developed by 15 people x 10 years
  - 1 major versions per year (Currently SAS 10.0), 1-2 minor releases
  - Various OS platforms supported
    - MacOS 10.6.X, Ubuntu X, MacOS 10.5.X, Fedora 12, RedHat 5, RedHat 4, Suse 11

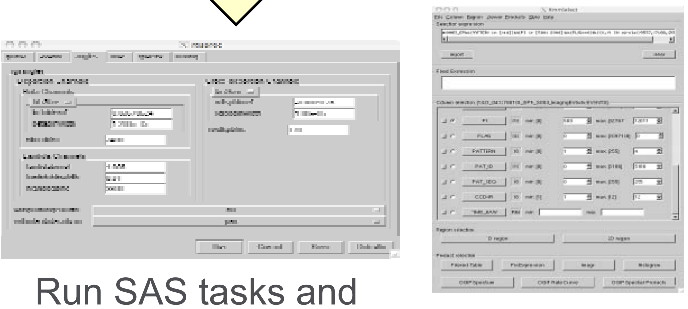
Traditional XMM-Newton SAS Workflow (SAS)

esa

ESAC

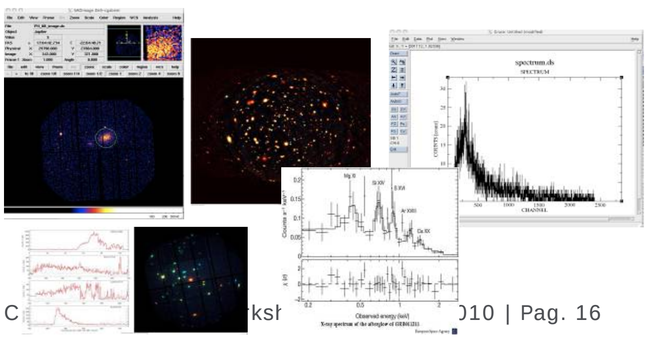
Download and Install SAS,
Calibration DB,
3rd party SW, libraries, …

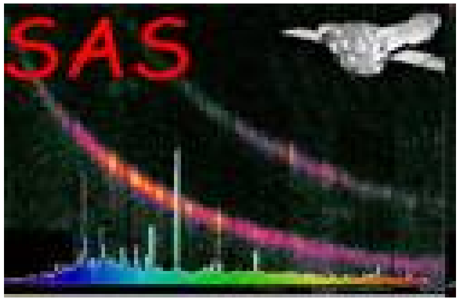
Search and download data
from XMM-Newton Archive



Run SAS tasks and
Store results
On Local computer

Vizualize data





European Space Agency

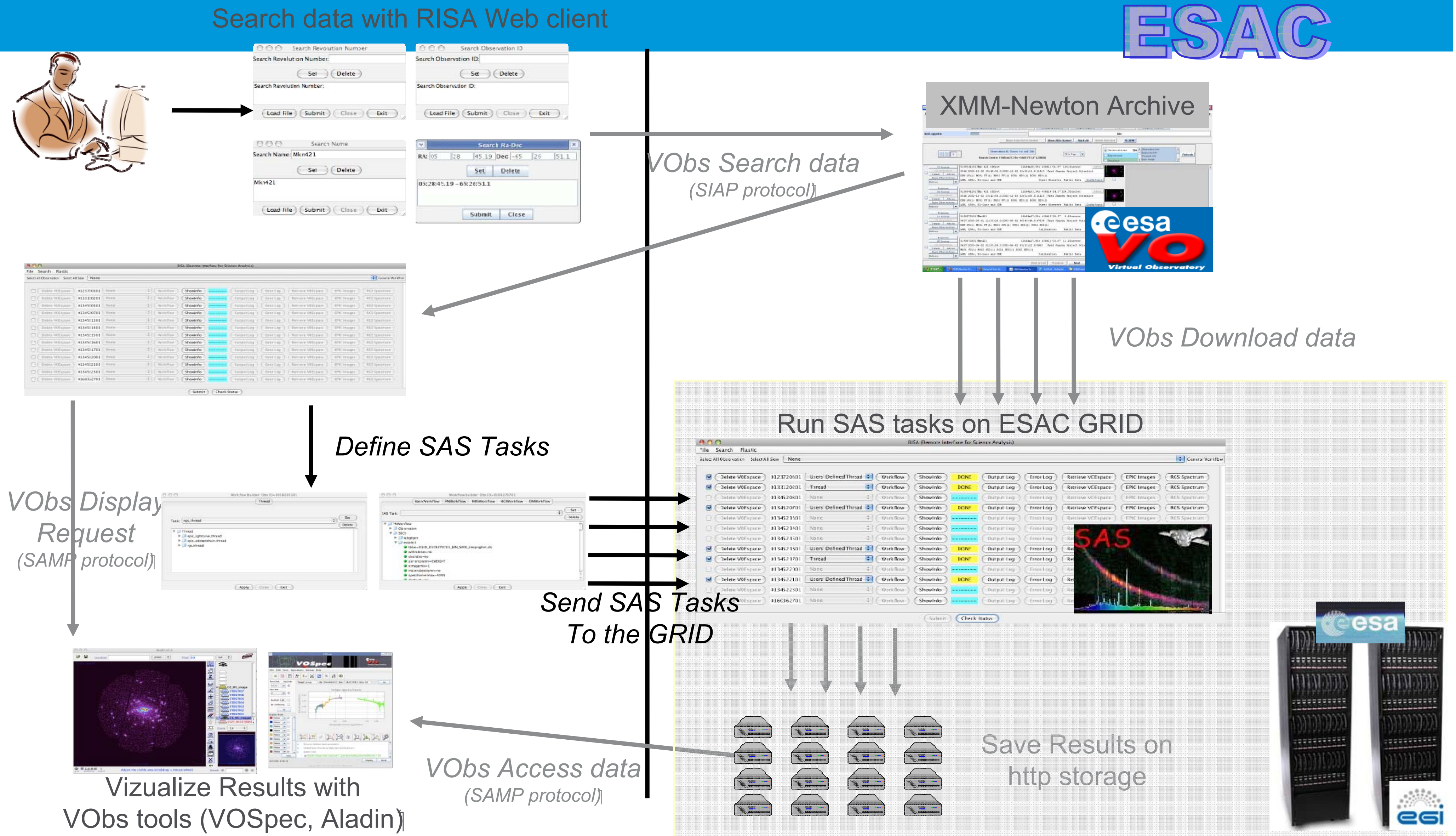
Towards a new way to run XMM-Newton SAS

- Run XMM-Newton SAS as a web services on the GRID
  - No need for the user to download and install SAS on his/her desktop
  - Automatic access to large HW and SW resources (ESAC Grid)
  - Access and process the data close to where they are (XMM-Newton Science Archive at ESAC, close to the ESAC Grid)
  - Usage of Virtual Observatory (VO) standards and tools

- SAS easier to use for the end users

- Potential reduction of SAS SW maintenance effort
  - Support required only to fewer platforms

XMM-Newton RISA –
Remote Interface for Science Analysis



European Space Agency

XMM-Newton Calibration Analysis - 1

- XMM- Newton EPIC-pn X-Ray imager Long-Term CTI (Charge Transfer Inefficiency) Analysis
  - The CTI is an important instrumental characteristic for event energy reconstruction. It increases during the instrument life time, basically due to accumulated radiation damage.
  - Careful monitoring since launch (1999) up till present is required
    - Build a calibration model of source line energies

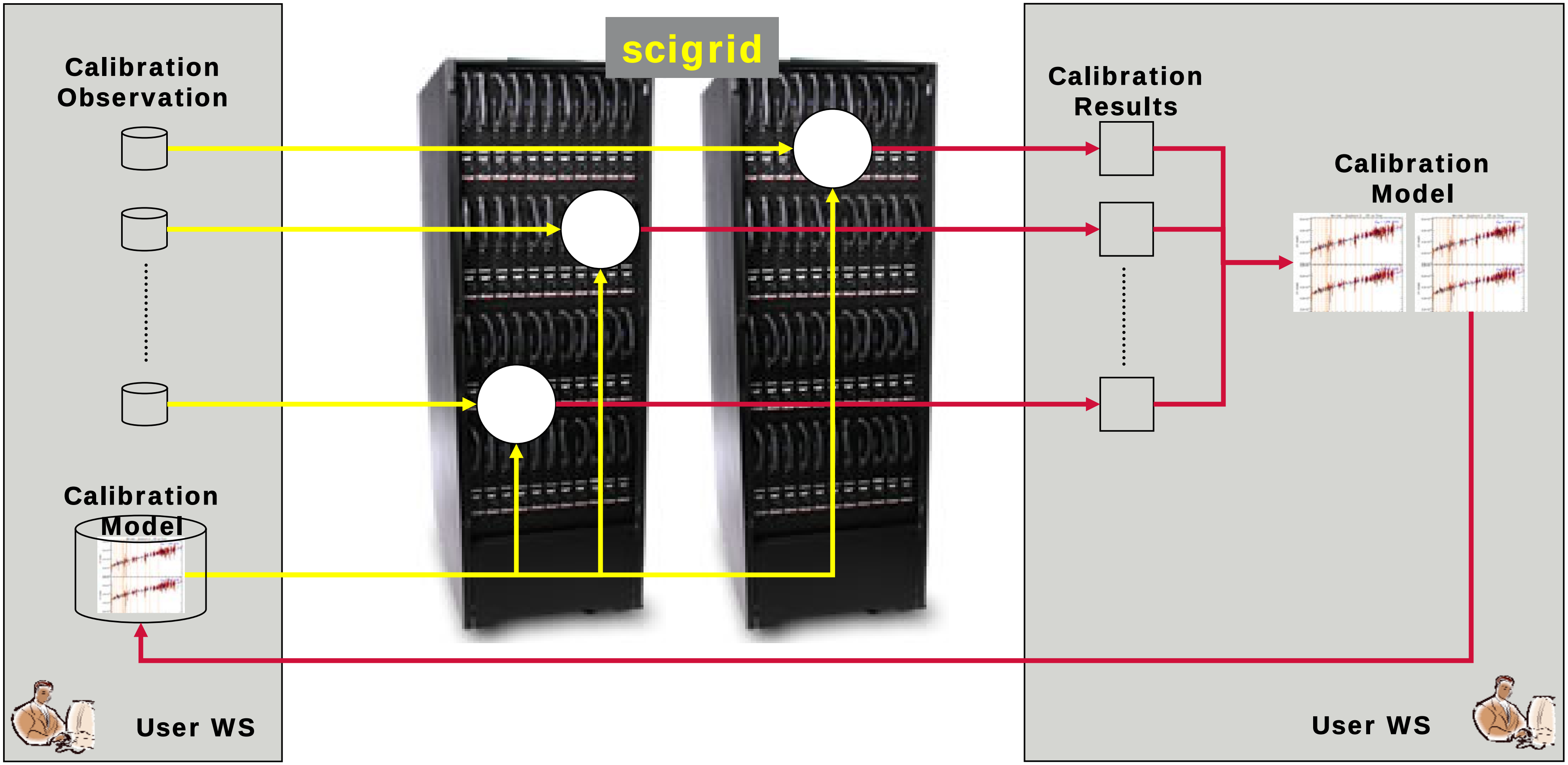
- Analysis of ~900 calibrations observations
  - ~7.2 Ms or 83 days of observing time
  - Parameters are derived and tweaked through several iterations of the whole process

- Each dataset processing : ~3hrs
- 900 datasets processing : 900 x 3 = 112.5 days = 3.75 months

- Using the GRID (full capacity)
  - (900 x 3) / 160 processors < 1 day !!!

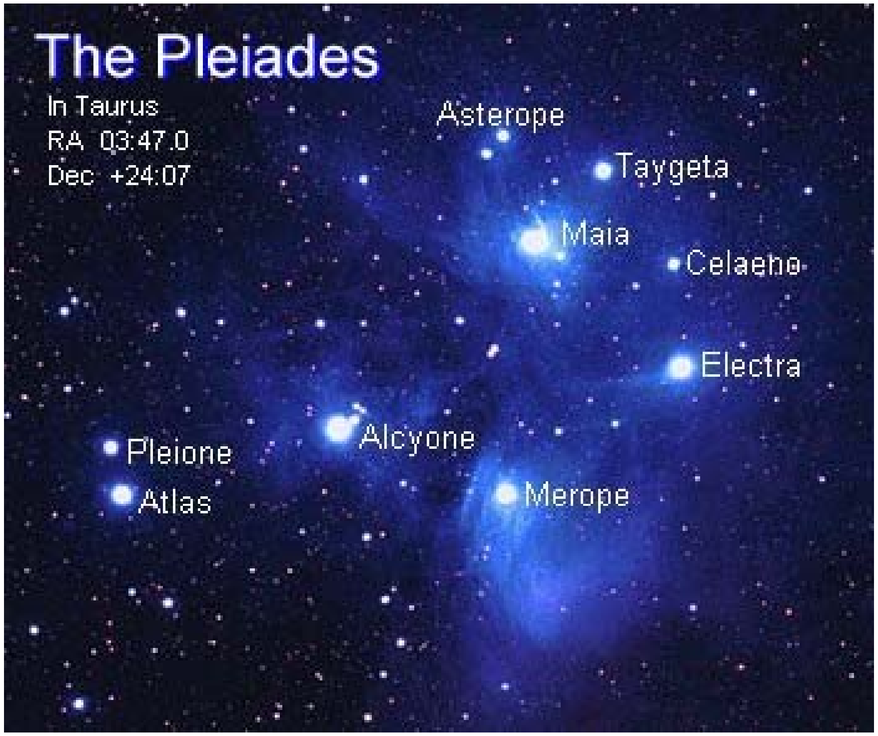
- GRID setup allows many iterations, hence making the whole exercise possible !

XMM-Newton Calibration Analysis - 2



Measure transverse motion of 730.000 stars - 1

- How many stars are in the Pleiades besides the famous ones?

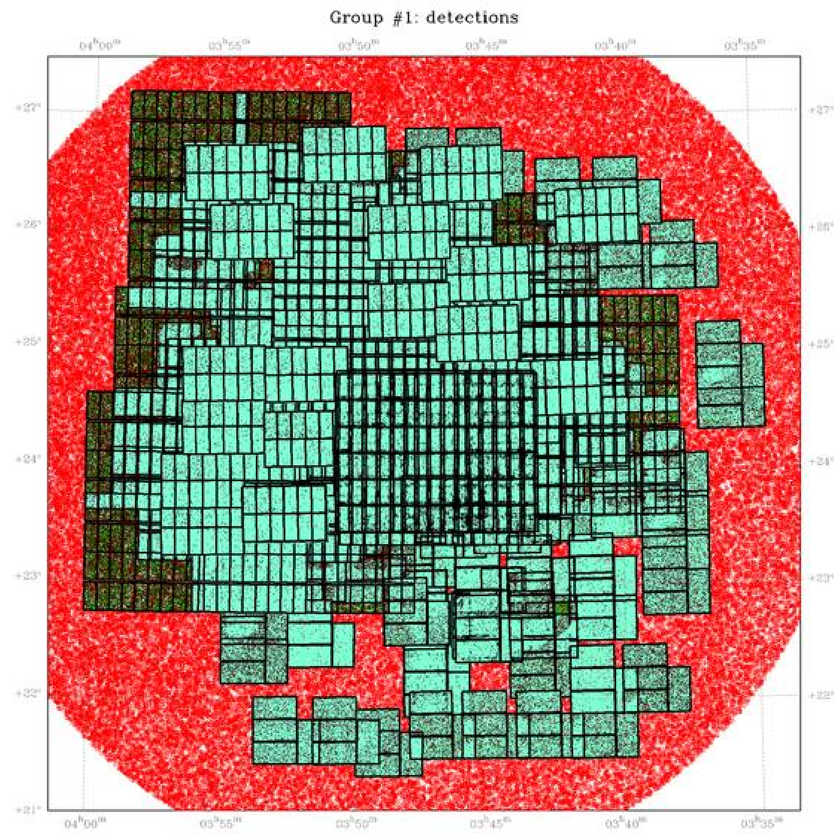


- Difficult to tell just from the right image
- The only good way to find out is to look for common transverse motion

Measure transverse motion of 730.000 stars - 2

- Use of Scigrid to process 520Gb of wide field images obtained from various space and ground based telescopes (Subaru, HST, CFHT, Isaac Newton Telescope) over a 12 years period
- The multi-epoch images were used to derive the transverse motion of every star (~730.000 in total) present in the 6deg x 6deg field of view of our observations

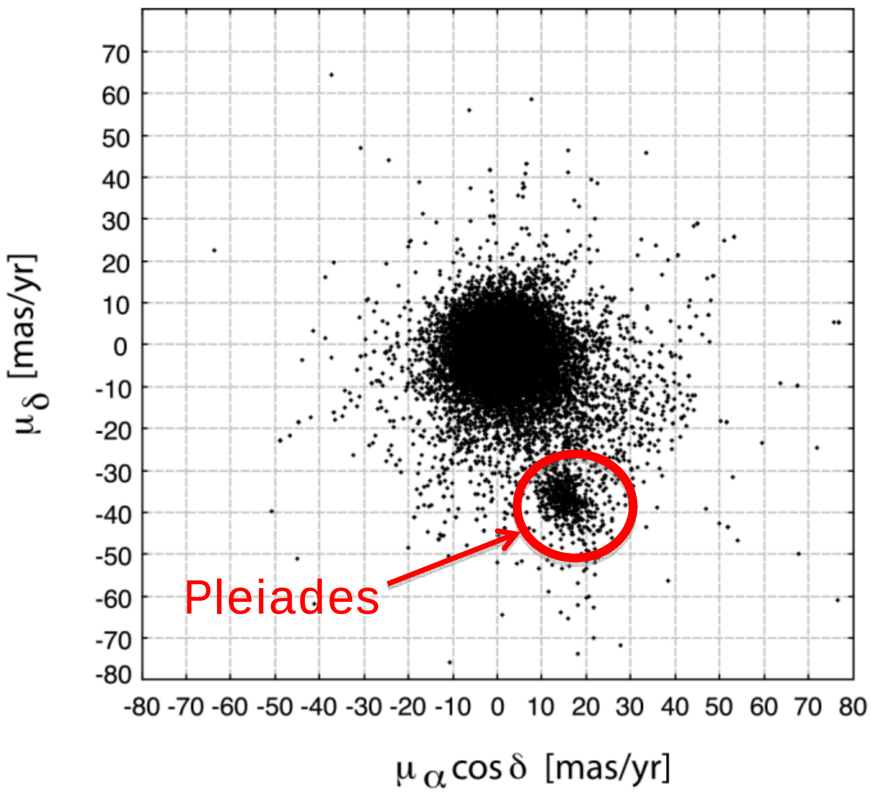
- The usage of Scigrid was critical, as it requires:
  - vast amount of storage
  - fast multi-threaded computers to extract the source photometry and astrometry
  - fast multi-threaded computers and vast amounts of RAM to cross-match
  - the multi-epoch catalogs and derive the kinematics of each star



H.Bouy et al

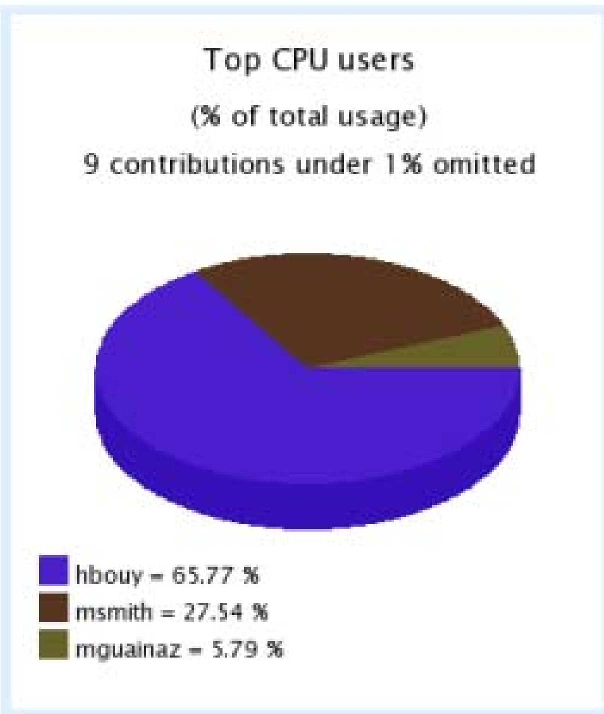
Measure transverse motion of 730.000 stars - 3

- The result is a beautiful vector point diagram of the motion of all 730.000 located in the Pleiades cluster. The diagram shows the Field and background objects are distributed randomly around (0,0) mas/yr, while the Pleiades members are all co-moving and form the locus near (15,-35)
- This allowed to unambiguously identify several thousands of members (when less than 1000 were known to date) down to the planetary mass regime.
- The scientific outcome will be extremely valuable and rich, from refining the mass function of the cluster, identifying planetary mass objects, and detailed studies of the internal dynamics.
- This new technique could be repeated for other regions of the sky



motion in RA (x-axis) and Dec (y-axis)
H.Bouy et al

Scigrid usage statistics from Jun-Sept 2010



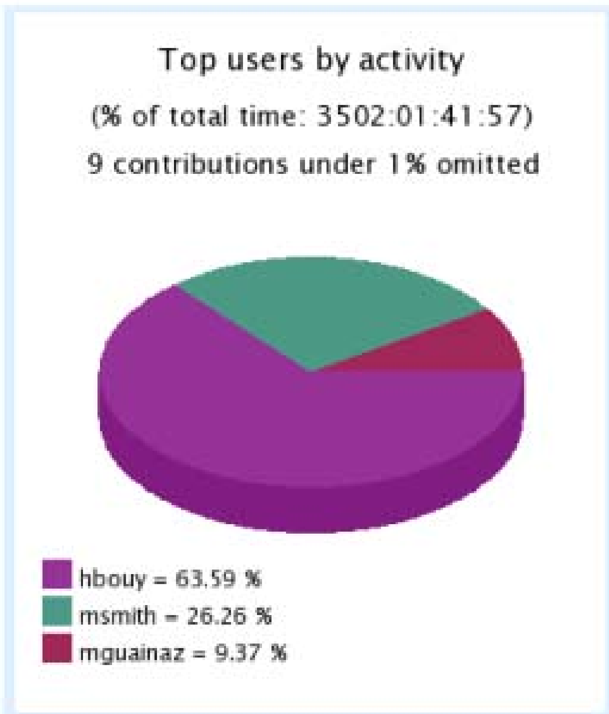
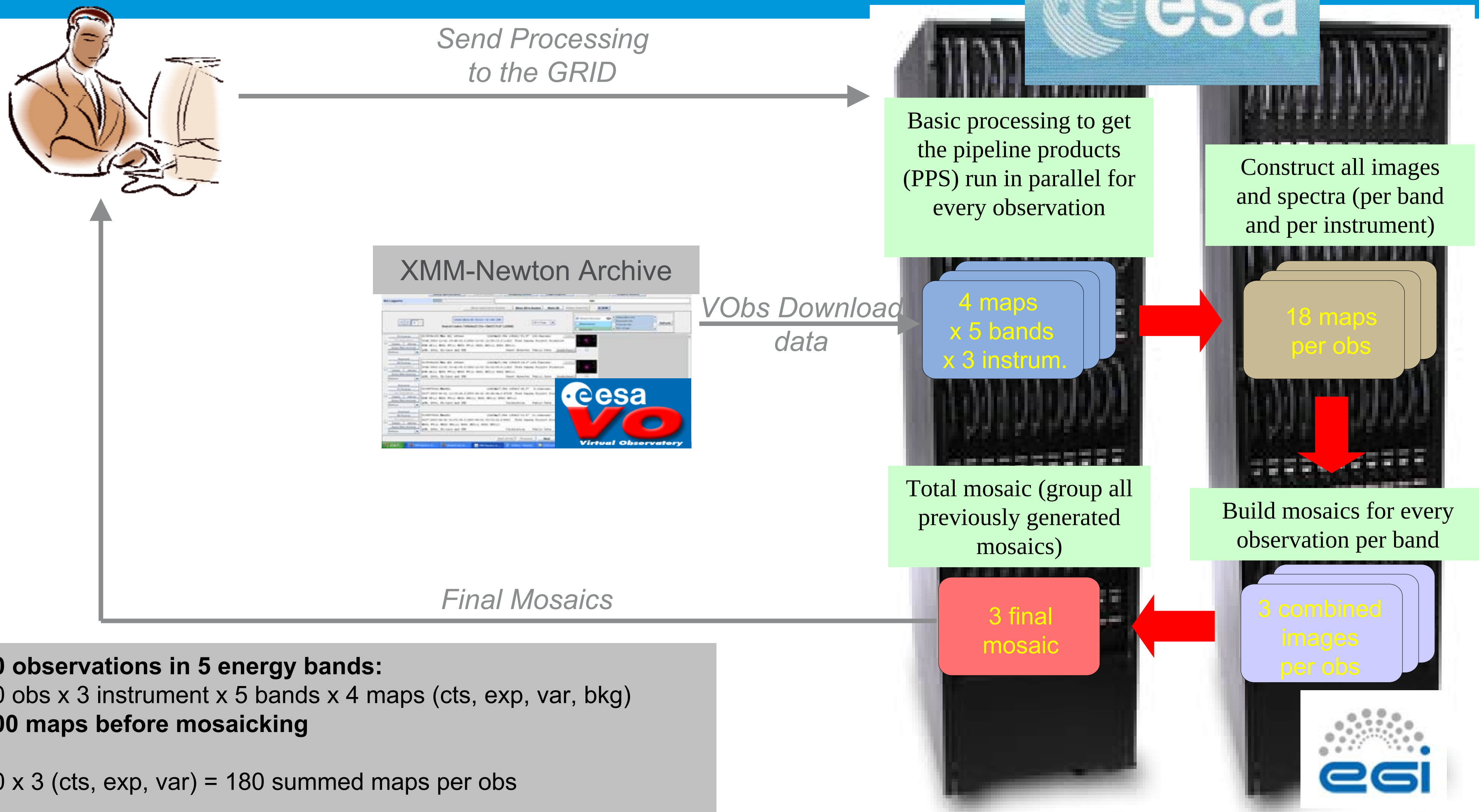


Table sorted by CPU usage time. To sort the columns, please click on the headers.

Rank	User	Total jobs	Total execution time	CPU time	RAM usage	Host
1	hbouy	29289 (33.95% of all jobs)	2226:22:10:41	2090:05:33:08	56.140825 TB	scigrid
2	msmith	25851 (29.97% of all jobs)	919:14:42:08	875:10:42:31	11.226931 TB	scigrid
3	mguainaz	29212 (33.86% of all jobs)	328:01:16:40	184:02:36:57	10.709227 TB	scigrid
4	icalle	645 (0.75% of all jobs)	8:03:39:15	8:01:22:05	471.514 GB	scigrid
5	jvallejo	3 (0% of all jobs)	7:11:00:56	7:10:41:47	19.742 GB	scigrid

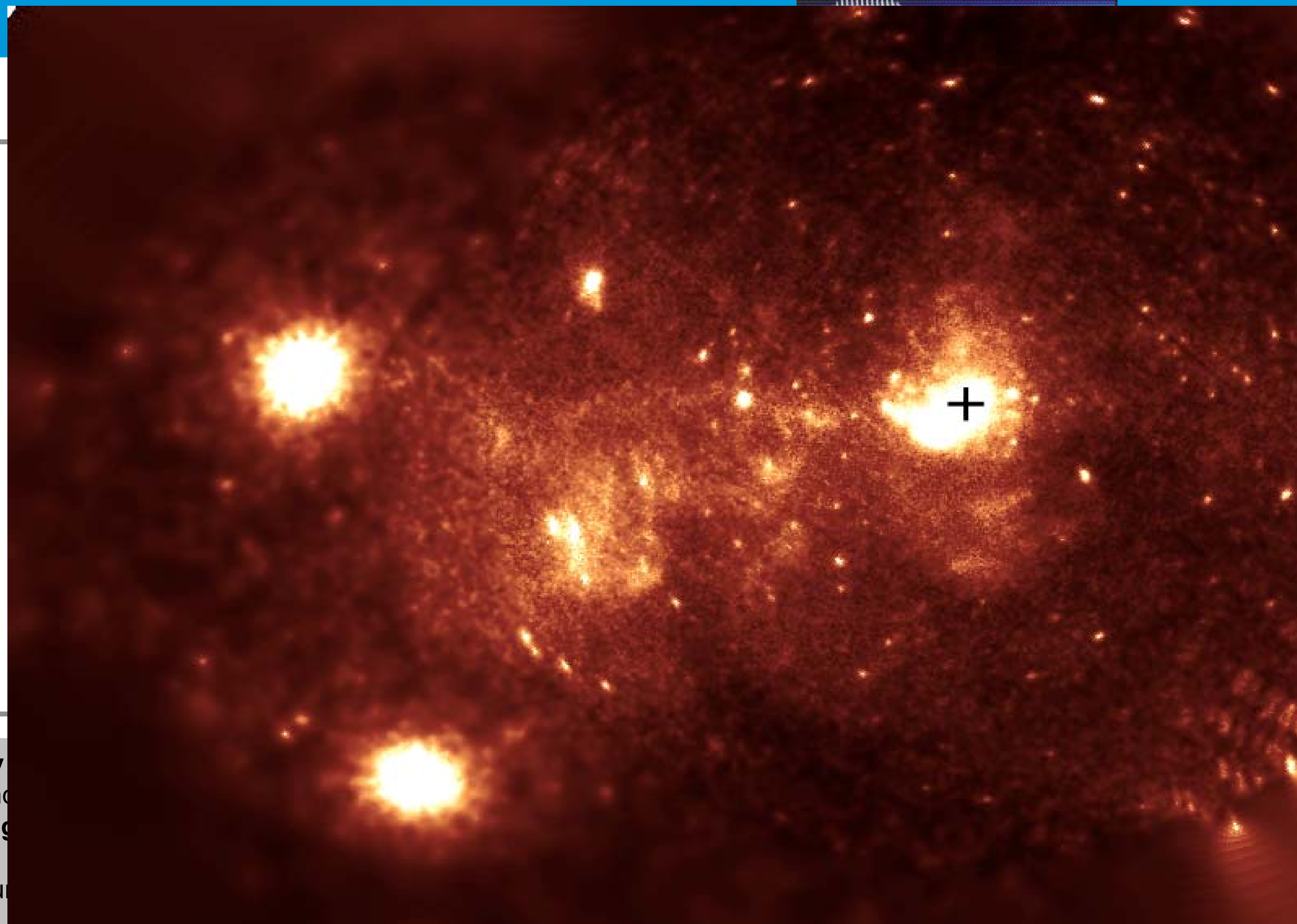
XMM-Newton Mosaic Construction



10 observations in 5 energy bands:
10 obs x 3 instrument x 5 bands x 4 maps (cts, exp, var, bkg)
600 maps before mosaicking
+
60 x 3 (cts, exp, var) = 180 summed maps per obs

10 obs = 780 maps

XMM-Newton Mosaic Construction



10 observations in 5 energy
10 obs x 3 instrument x 5 band
600 maps before mosaicking
+
60 x 3 (cts, exp, var) = 180 su

10 obs = 780 maps

Cloud and GRID : Future Directions

- Investigating Cloud concepts, while keeping the existing infrastructure
  - Cloud and GRID co-existence, internally and externally

1. Deploy an internal private cloud at ESAC
   - Allow easier and more flexible redeployment of hardware resources (for GRID and non GRID usage)
   - Deploying on-demand ESAC GRID worker nodes on internal private cloud at ESAC when more Grid resources are required (build on existing GRID infrastructure)

2. Use public cloud to access more hardware resources
   - Deploy complete ESAC applications on public cloud (eg Gaia proof of concept)
   - Deploying on-demand extra worker nodes on public (eg Amazon) cloud using a VPN (only valid for ESAC GRID, not SCIGRID)
   - Need to pay on demand

Acknowledgments

- ESAC Computer Support Group
- ESAC Science Archives and Virtual Observatory Team
- ESAC Herschel Data Processing Team
- ESAC XMM-Newton SAS Team
- Michael Smith for the XMM-Newton Calibration monitoring
- Bruno Merin, Herve Bouy and Guillaume Belanger for their science work