



# Science Data Management at ESO

Martino Romaniello  
Head, Back-end Operations Department  
European Southern Observatory





# Ensuring the science value of data, enabling its exploitation

- Providing the best science data is at the core of ESO's mission to enable major science discoveries from our science community
  - A community that is very large in number and varied both in interests and in skills
- So, for the **La Silla Paranal Observatory** we
  - 1. Ensure that instruments are working properly
  - 2. Ensure that the science content can be extracted from the data
  - 3. Deliver the science data to our users, PIs and archive researchers alike
- I will introduce and put in context these different activities, which will be expanded upon during the Workshop

→ **ALMA: Stoeck, Wednesday at 16:00**



# Act 1.

**Ensuring that instruments are  
working properly:  
the Quality Control loop**



# The Quality Control loop


- The current performance of the instruments is constantly measured in a quality control process
  - Quality and compliance w/ user constraints is first evaluated online
  - Raw data is transferred in real time to Headquarters and processed into products
  - Relevant health-check parameters are, then, measured, trended and the results are fed back to the Observatory for immediate follow up, as needed
  - Calibration completeness and quality is checked
  - This detailed knowledge of the instrumental (and atmospheric) signatures is, then, also crucial to reveal the science signal

→ Mieske and Wolff, Wednesday at 11:00  
→ Percheron, Wednesday at 16:55  
→ Posters : Dobrzycka, Hummel





# Information on demand



mirror sites: [PL](#) (internal link) [HQ](#) [?](#)

AL | HC | refs | QC

QUALITY CONTROL

HOME

UT1

- FORS2
- KMOS
- NACO

UT2

- FLAMES/GIRAFFE
- UVES&FLAMES/UVES
- X-SHOOTER

UT3

- SPHERE
- VIMOS
- VISIR

UT4

- HAWK-I
- MUSE
- SINFONI

VLT1

- AMBER

Survey Cameras

- OMEGACAM
- VIRCAM

General QC:

- [Image Quality \(historical\)](#)
- [Sky Background \(historical\)](#)

General process:

- [pre-imaging](#)

Retired:

- FORS1 (April 2009)
- ISAAC (December 2013)
- CRIRES (July 2014)
- MIDI (March 2015)

Related:

- Instruments
- Paranal SciOps
- User Support
- Archive
- Exposure Time Calc.
- ESO acronyms

QC links:

- QC home
- Data Products
- calChecker
- HealthChecks
- Reference Frames

QC1 database

- Paranal autrep database (ESO internal)

## Quality Control and Data Processing

This is the home page of the Quality Control and Data Processing Group (QCG) within ESO's Data Management and Operations Division (DMO). Find here information about the pipeline-processing of data from the VLT and about their quality control.

QuickLinks:

Trending & QC1

Pipelines

Data packages (until 2011-09 only)

VLT instruments:

Go

VLT instruments:

Go

VLT instruments:

Go

At ESO Headquarters in Garching, calibration data from all Paranal instruments (VLT, VLTI, survey cameras) are processed, their quality is checked and monitored. This is the task of the **Quality Control and Data Processing Group** ("QC Garching"). This process is part of the back-end of the Data Flow System.

QC Garching has the following main tasks:

- ▶ monitor the calibration plan and report issues back to Paranal Observatory
- ▶ create master calibration data, control their quality and certify them
- ▶ perform instrument trend analysis and feed back the results to Paranal Observatory
- ▶ deliver master calibration data to the archive
- ▶ process [pre-imaging](#) data
- ▶ create science-grade data products for selected instruments and instrument modes.

Note:

- ▶ Pls can access their raw data and all associated calibration data on the [ESO User Portal](#).
- ▶ After the proprietary period has expired, all archive users can access these data.

[www.eso.org/HC](#)

HealthCheck Monitor: Information about health check parameters (daily refreshed, includes most recent nights)

[www.eso.org/CAL](#)

calChecker: calibration completeness monitor (refreshed every half hour)

Reference Frames

reference frames archive, for Paranal operational specialists' use

QC1 Database

download and plot QC1 parameters

▶ Science-grade data products:

UVES science-grade data products

UVES ECHELLE data, point sources: quality-controlled, homogeneous processing of complete data stream 2000-now

XSHOOTER science-grade data products

XSHOOTER SLIT data: quality-controlled, homogeneous processing of complete data stream 2009-now


GIRAFFE science-grade data products

GIRAFFE Medusa1/2 data: quality-controlled, homogeneous processing of complete data stream 2003-now

▶ Find [here](#) general information about the VLT and VLTI pipelines.

▶ [QC related publications](#) ...

# Information on demand



mirror sites: [PL](#) (internal link) [HQ](#) [?](#)

## Quality Control and Data Processing

[EAL](#) | [IHC](#) | [refs](#) | [QC](#)

**QUALITY CONTROL**

**HOME**

This is the home page of the Quality Control and Data Processing Group (QCG) within ESO's Data Management and Operations Division (DMO). Find here information about the pipeline-processing of data from the VLT and about their quality control.

**QuickLinks:**

[Trending & QC1](#)

**Pipelines**

**Data packages**  
(until 2011-09 only)

## MUSE trending system: overview of scores

Last update: 2015-11-23T16:08:52 (UT)

### General news:

## MUSE news:

**MUSE instrument score:**

▶ [red score comments ...](#) | [all comments ...](#) ▶ [edit COMMENTS ...](#)

HC navigation group	group score	report scores									
detector:bias	■	<a href="#">RON_master</a>	■	<a href="#">RON_det</a>	<a href="#">BIAS_median</a>	<a href="#">temp</a>	■				
detector:dark	■	<a href="#">DARK_med</a>	■								
flats/lamps	■	<a href="#">FLAT_mean</a>	■	<a href="#">FLAT_algnm</a>	<a href="#">TRACE_lxpos</a>	<a href="#">TRACE_rxpos</a>	<a href="#">TRACE_tilt</a>	<a href="#">TRACE_width</a>	<a href="#">TRACE_vignet</a>	■	
wave	■	<a href="#">ARC_wavecal-N</a>	■	<a href="#">ARC_wavecal-E</a>	<a href="#">RESOL_det</a>	<a href="#">ARC_fit-N</a>	<a href="#">ARC_fit-E</a>	<a href="#">ARC_flux</a>	■		
throughput	■	<a href="#">THROUGHPUT</a>	■								
detmon	■	<a href="#">gain</a>	■	<a href="#">gain_det</a>	■						

### ESU acronyms

**QC links:**

[QC home](#)

## Data Products

## calChecker

**HealthChecks**

Reference Fram

### REVISIONS

QC1 database

[Paranal autrep database](#)

(ESO internal)

**GIRAFFE science-grade data products**

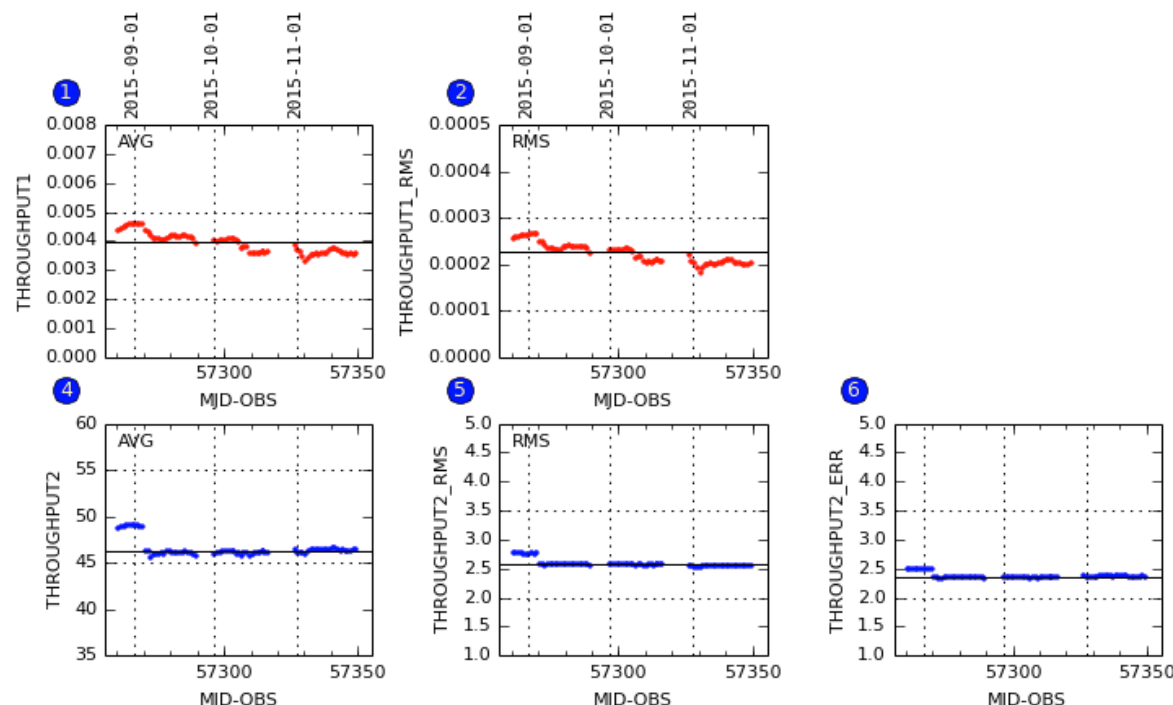
GIRAFFE Medusa1/2 data: quality-controlled, homogeneous processing of complete data stream 2003-now

► Find [here](#) general information about the VLT and VLTI pipelines.

► [QC related publications ...](#)

# Information on demand

MUSE: THROUGHPUT (last 90 days)  
QC data range: 2015-08-26 ... 2015-11-22\*



## Quality Control and Data Processing

in ESO's Data Management and Operations Division (DMO). Find here  
by control.

Data packages  
(until 2011-09 only)

**cores**

temp

TRACE rxpos ■ TRACE tilt ■ TRACE width ■ TRACE vignet

ARC\_fit-N ■ ARC\_fit-E ■ ARC\_flux ■

wave
throughput
detmon

<u>ARC_wavecal-</u>	<u>ARC_wavecal-</u>
<u>N</u>	<u>E</u>
<u>THROUGHPUT</u>	
<u>gain</u>	<u>gain_det</u>

ESU aClonvins

2C links:

[QC home](#)

## Data Products

calChecker

CalChex  
HealthChecks

## Reference Frames

## Reference List

QC1 database

Paranal autrep database

Paranal autre p  
(ESO internal

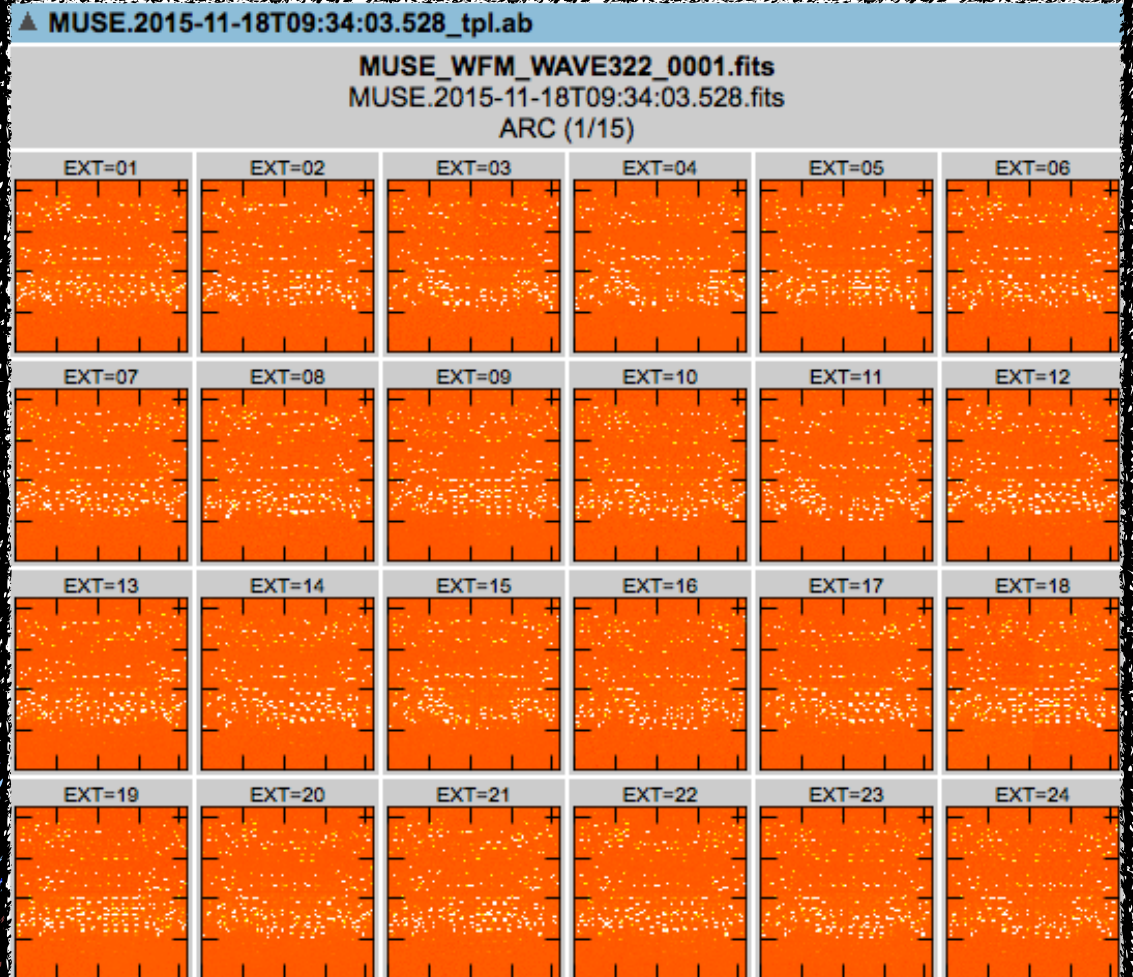
**GIRAFFE science-grade data products**

GIRAFFE Medusa1/2 data: quality-controlled, homogeneous processing of complete data stream 2003-now

► Find [here](#) general information about the VLT and VLTI pipelines.

► [OC related publications](#) ...

## Quality Control and Data Processing



**QC links:**  
[QC home](#)  
[Data Products](#)  
[calChecker](#)  
[HealthChecks](#)  
[Reference Frames](#)

[QC1 database](#)  
[Paranal autopilot database](#)  
 (ESO internal)

- Find [here](#) general information about the VLT and VLTI pipelines.
- [QC related publications](#) ...





# Information on demand

MUSE: THROUGHPUT (last 90 days)  
QC data range: 2015-08-26 ... 2015-11-22\*

Quality Control and Data Processing

▲ MUSE.2015-11-18T09:34:03.528\_tpl.ab

MUSE\_WFM\_WAVE322\_0001.fits

MUSE.2015-11-18T09:34:03.528.fits

ARC (1/15)

EXT=01

EXT=02

EXT=03

EXT=04

EXT=05

EXT=06

THROUGHPUT

2015-11-17: MUSE

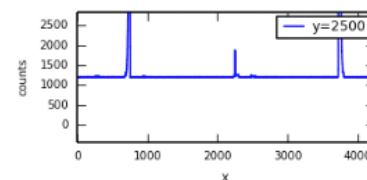
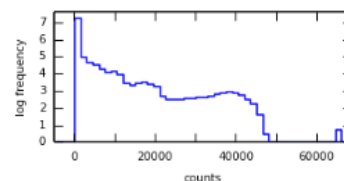
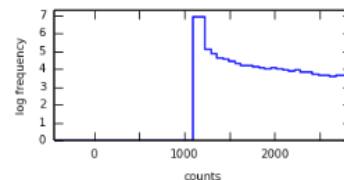
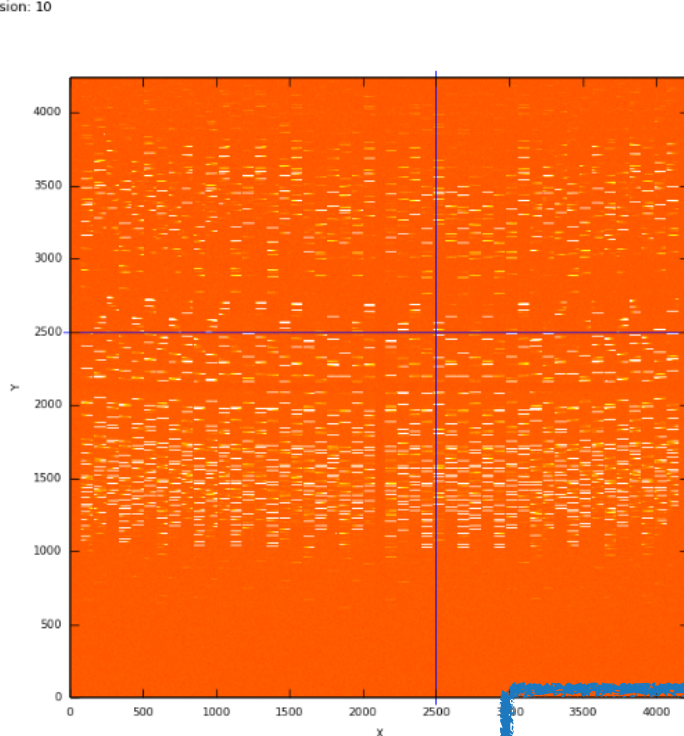
ORIGFILE: MUSE\_WFM\_WAVE322\_0001.fits  
ARCFILE: MUSE.2015-11-18T09:34:03.528.fits  
raw\_type: ARC  
data\_class: ARC  
extension: 10

type info

DPR.CATG: CALIB  
DPR.TYPE: WAVE  
DPR.TECH: IFU  
TPL.ID: MUSE\_wfm\_cal\_wave

set-up info

INS.MODE: WFM-NOAO-N  
INS.OPT11.NAME: Blue  
DET.READ.CURNAME: SC11.0



T=09

EXT=10

EXT=11

EXT=12

T=15

EXT=16

EXT=17

EXT=18

T=21

EXT=22

EXT=23

EXT=24

[www.eso.org/qc](http://www.eso.org/qc)



## **Act 2.**

**Ensuring that science can be  
extracted from the data:  
science data products**



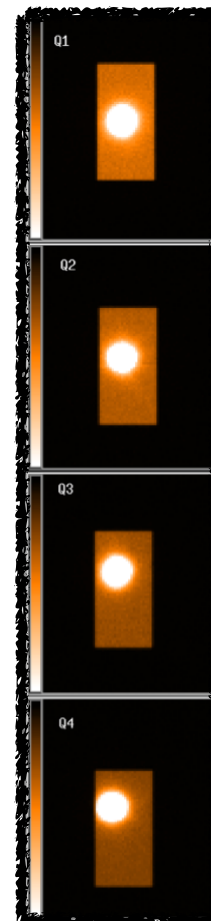
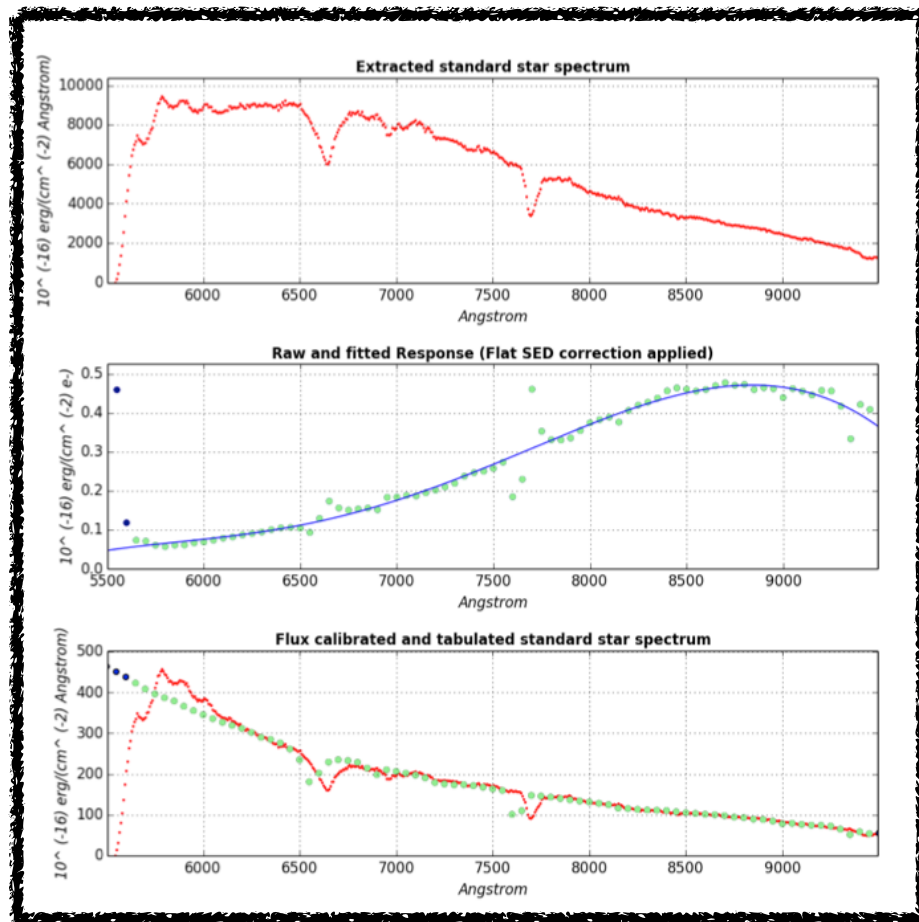
# Science data products - I

- Everything needs to be in place so that science can be extracted from the data
  - The appropriate observing procedures are in place
  - The appropriate calibrations are taken
  - Suitable data reduction tools (aka “pipelines”) are available
- Data reduction tools are in operation at the Observatory and at Headquarters and are made available to science users for desktop data reduction
  - Pipelines are available for all Paranal instruments, covering the large majority of instrument modes and virtually the entire data volume

→ Freudling, Thursday at 14:55  
→ Ivanov (poster)

# Science data products - II

- Data product quality is more than just quality pipelines...



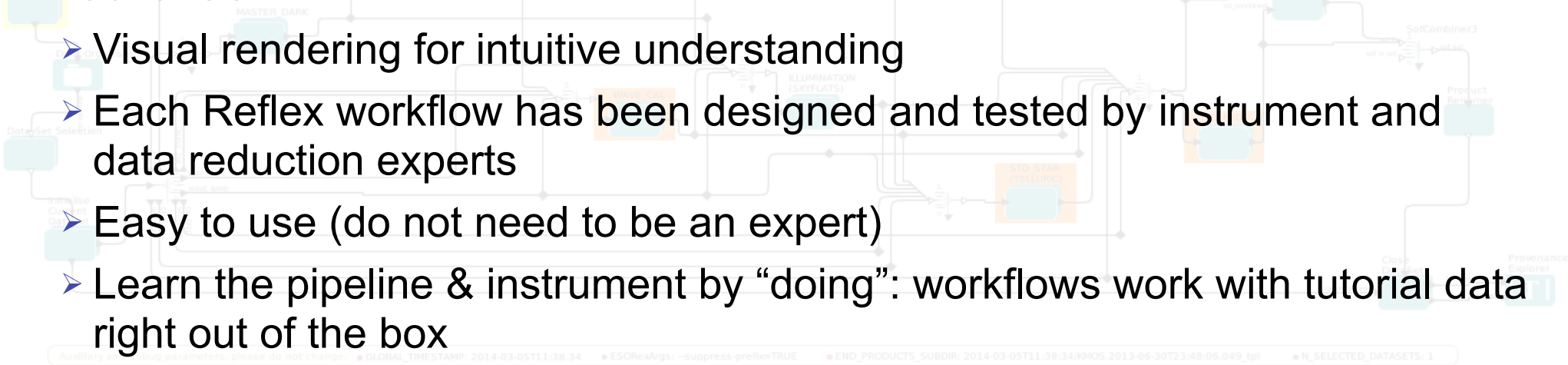
The cause of the wavelength shifts were the telescope offsets applied between quadrants



on science

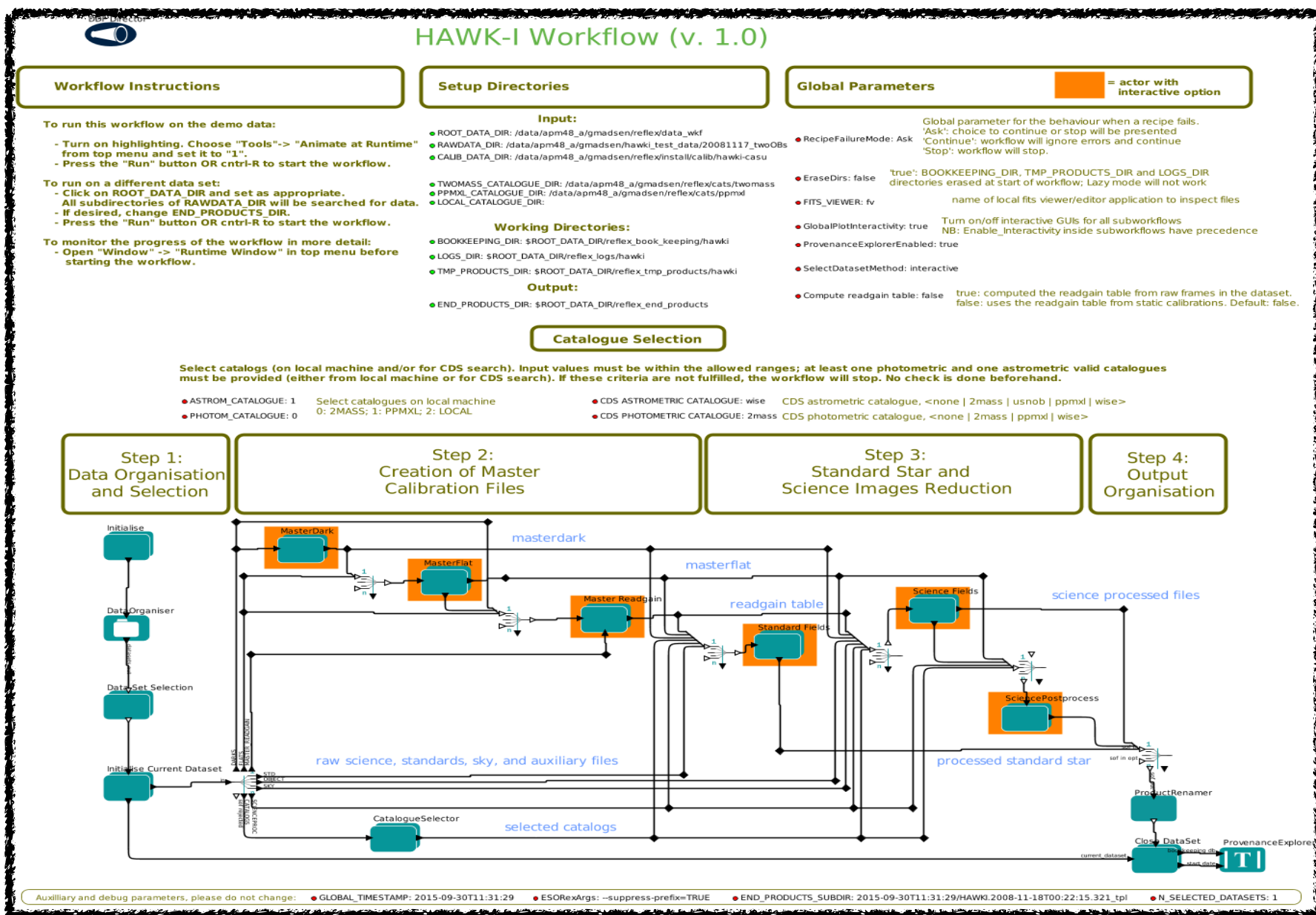
- Organizes the data for you
- Pipeline modules run with a single click
- You can monitor the progress of the workflow
- Does the book-keeping for you
- Allows user interaction and modification of processing parameters
- Allows insertion of user procedures in any language

## ■ Features

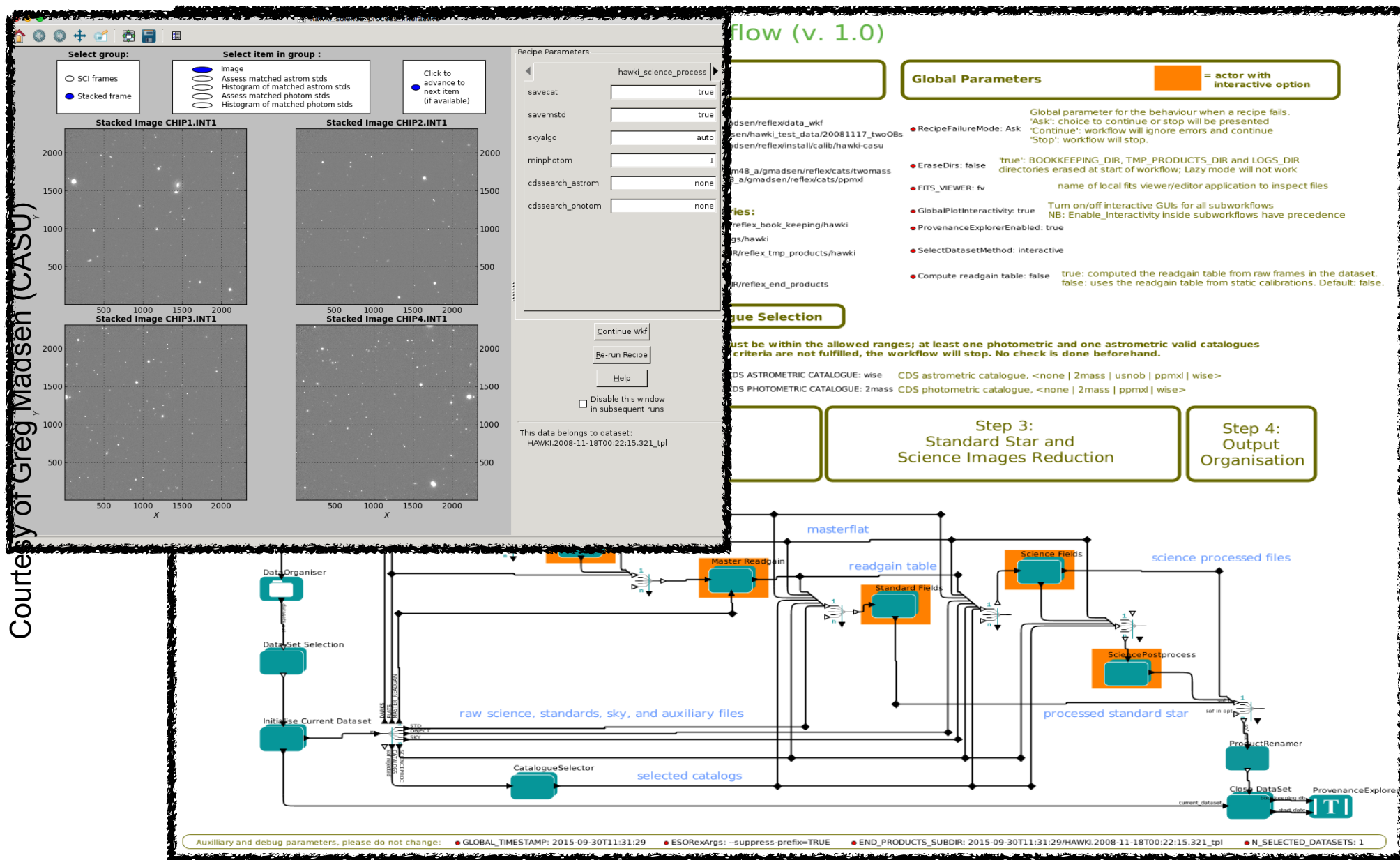
- 
- Visual rendering for intuitive understanding
  - Each Reflex workflow has been designed and tested by instrument and data reduction experts
  - Easy to use (do not need to be an expert)
  - Learn the pipeline & instrument by “doing”: workflows work with tutorial data right out of the box

# ESO Reflex

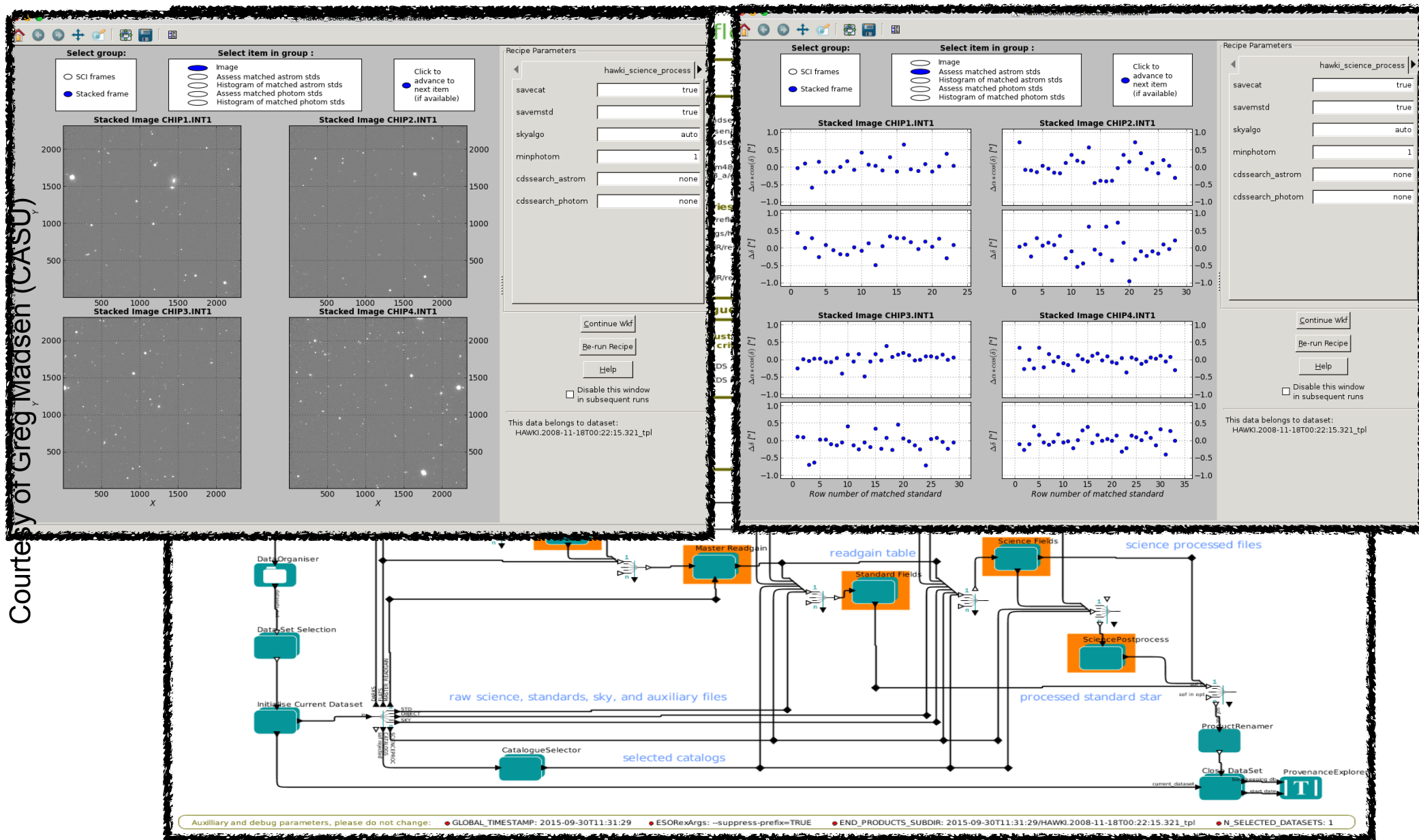
Courtesy of Greg Madsen (CASU)



# ESO Reflex



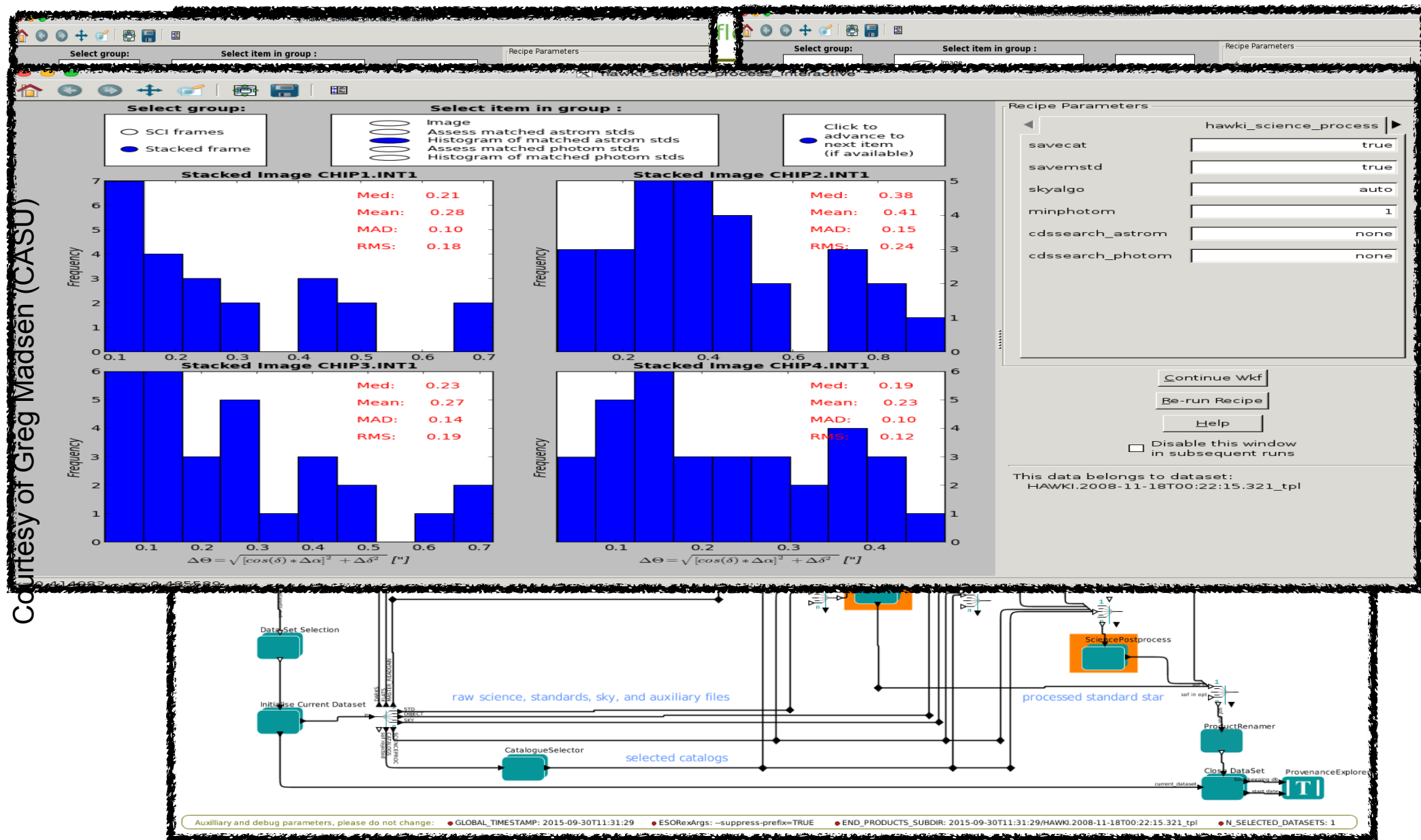
# ESO Reflex



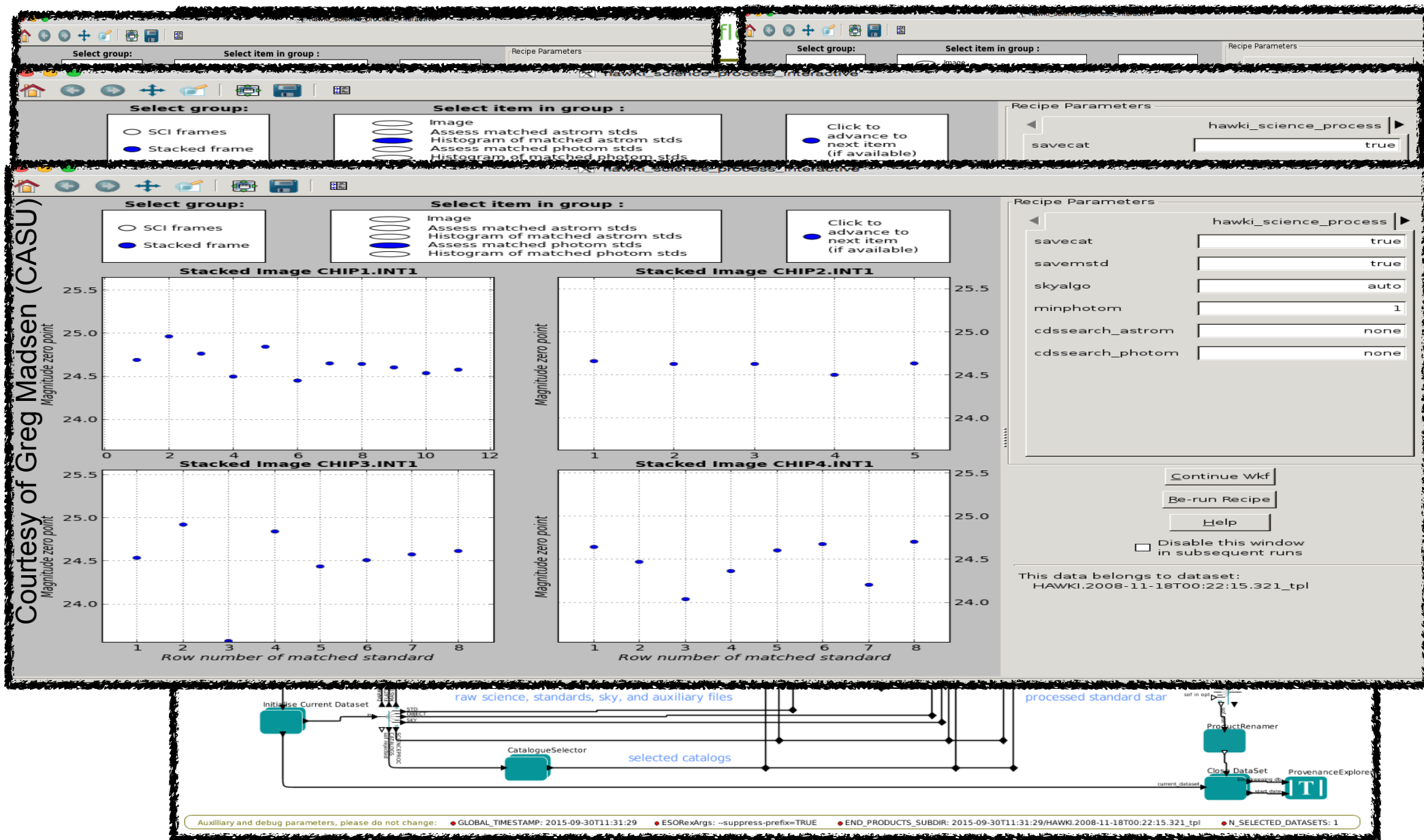


# ESO Reflex

Courtesy of Greg Madsen (CASU)

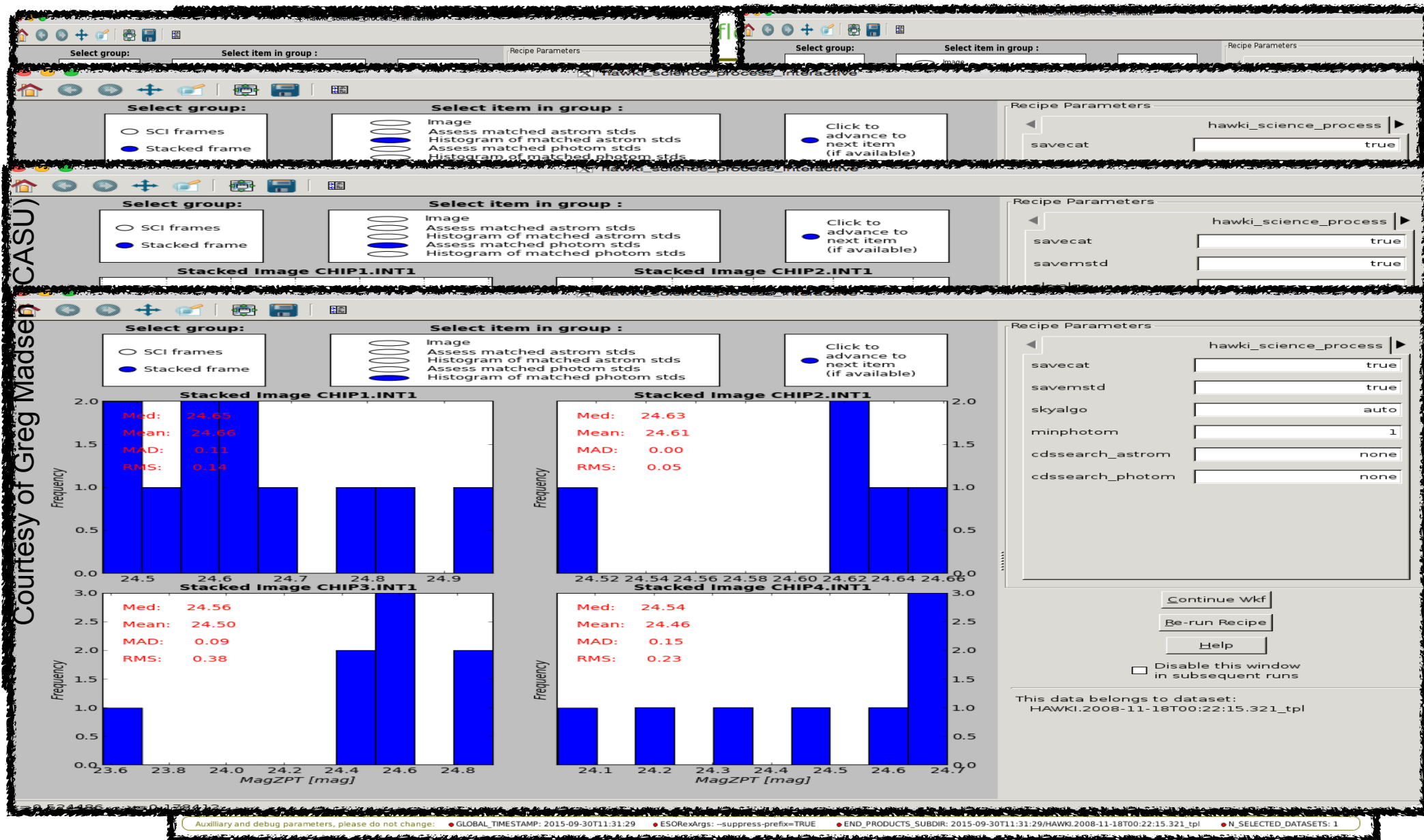


# ESO Reflex





# ESO Reflex





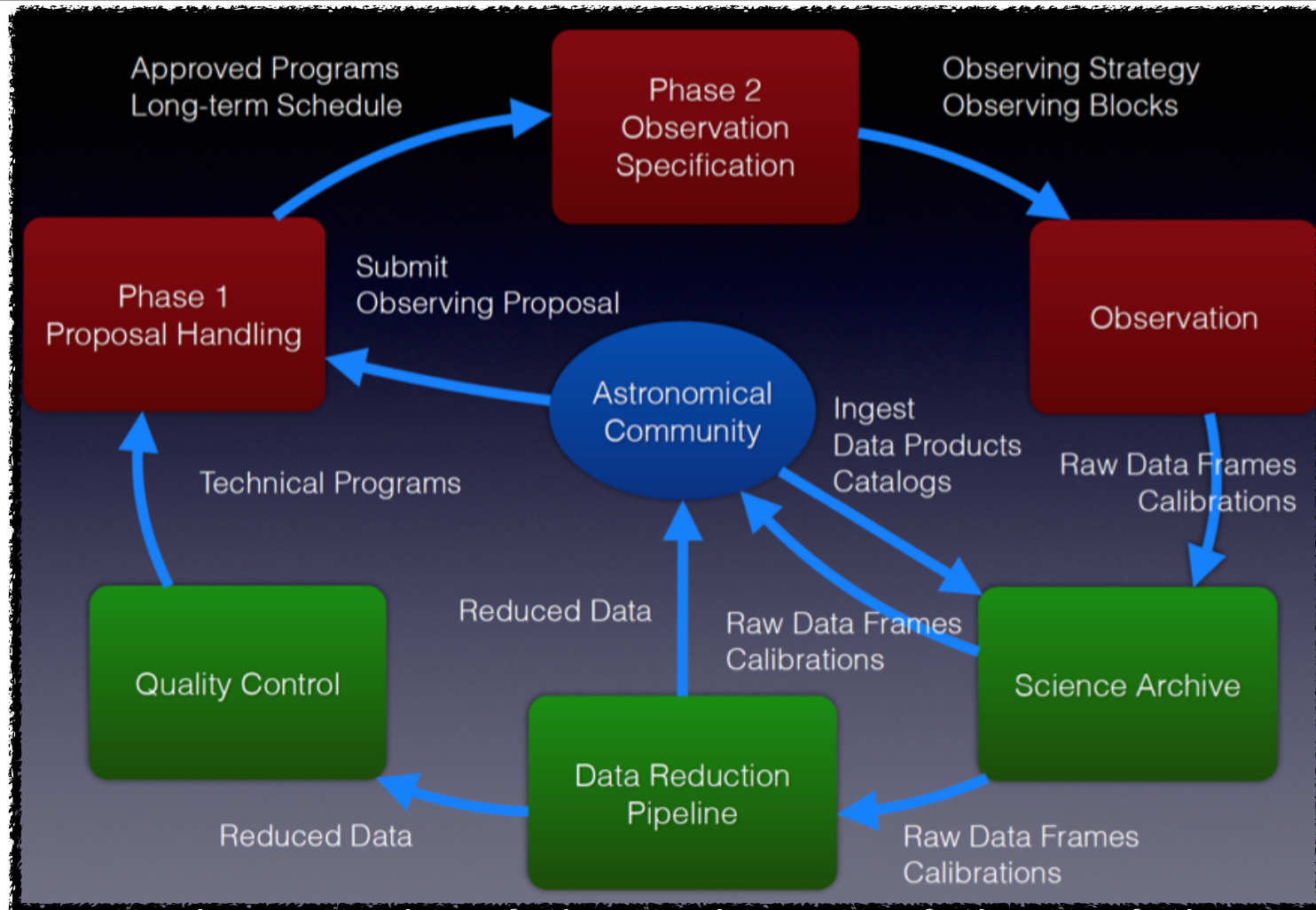
# **Interlude.**

**The context:  
end-to-end operations**



# The end-to-end operation model

Courtesy of Thomas Bierwirth



→ Ballester, Tuesday at 16:45  
→ Rejkuba, Wednesday at 10:10



## **Act 3.**

# **Delivering the science data: the ESO Science Archive Facility**





# Science Archive Facility cheat sheet

- The ESO Science Archive Facility (SAF) is the operational and technical data archive of the La Silla Paranal Observatory and a science resource in its own right
  - It is the one access point to La Silla Paranal data
- Data holdings (raw data and data products)
  - ~500 TB of data in 25 million files and ~23 billion database rows worth of header keywords
  - Storage technology is not a concern, database management more challenging
  - Inflow: ~12 TB/month; outflow ~15 TB/month

→ Fourniol (poster)

# Data products in the SAF

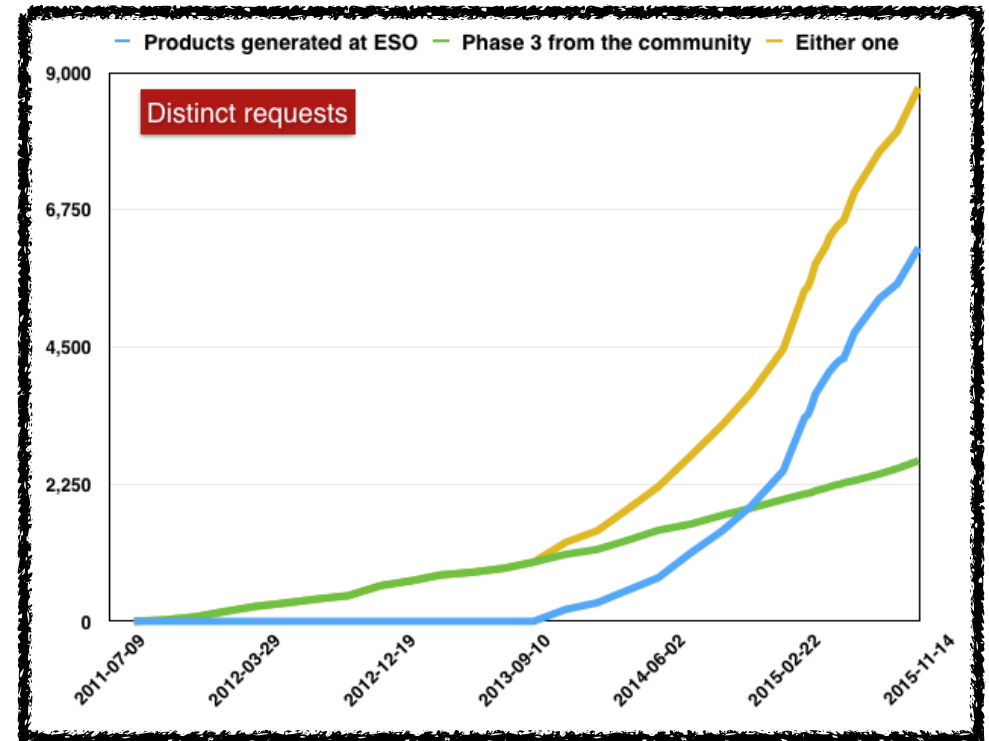
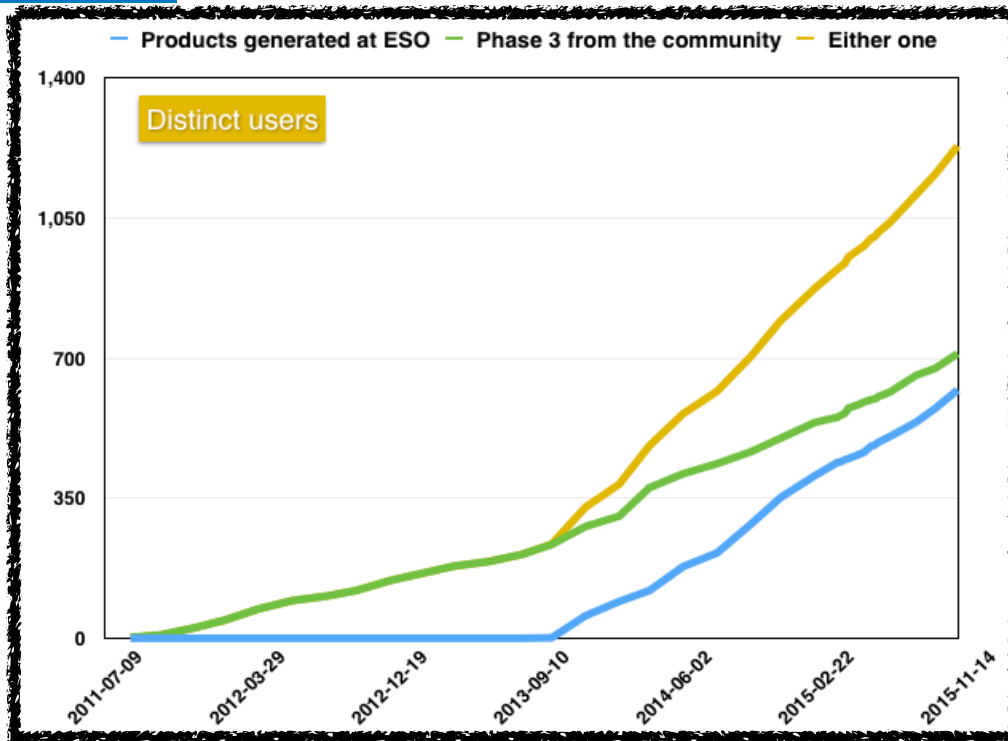
- Science ready data products through the Science Archive Facility to foster their quicker, wider use
- Two channels to feed the archive with data products
  - **Internal:** automated processing with scientifically validated pipelines
    - UVES echelle, X-Shooter echelle, HARPS echelle, FLAMES-MEDUSA, then HAWK-I and VIMOS IMG (UK in-kind), KMOS, MUSE, ...
    - Migration into Phase 3 of legacy historical Advanced Data Products completed
  - **External:** Principal Investigators of Public Surveys, Large Programmes, ... provide high-level products (mosaics, source catalogues, ...) that we validate and integrate
- Summary in December 2013 Messenger: papers from ESO, the Survey Teams and archive users
- Building high-quality, extensive content
  - [http://www.eso.org/sci/observing/phase3/data\\_releases.html](http://www.eso.org/sci/observing/phase3/data_releases.html)



Arnaboldi et al (2013)

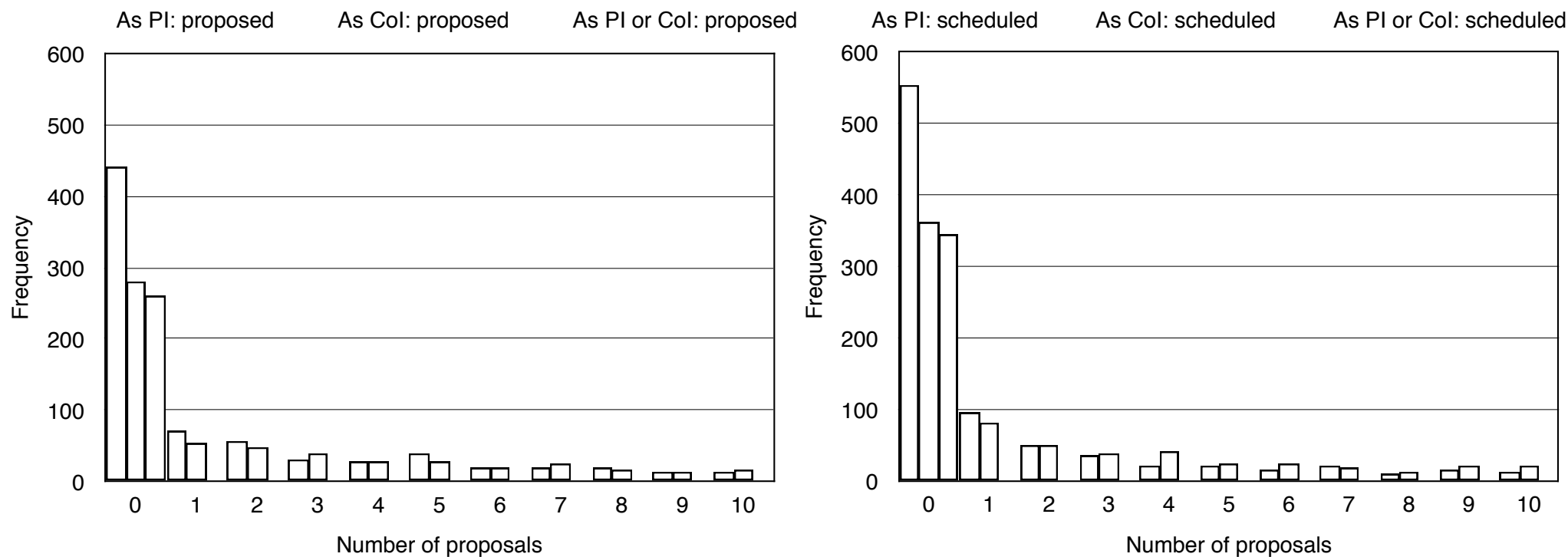
- ➔ Hanuschik, Wed 12:20; Retzlaff, Thu 17:15; Delmotte, Thu 17:35
- ➔ Mascetti (poster)

# Archive access to data: building a community

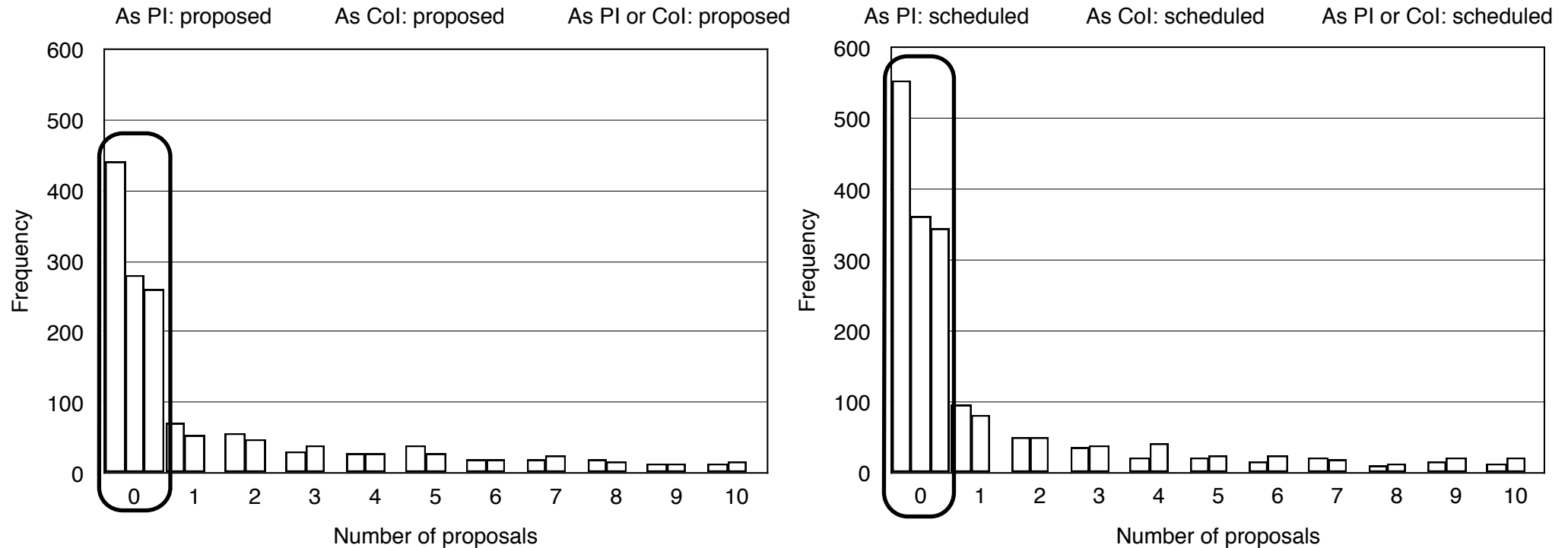


- Both pipeline data products generated at ESO and data products from the community are very much in demand
  - Archive users query for reduced products more than once
  - The number of archive users of reduced products is 1.5 times the number of PIs/Cols of Public Surveys
- Availability of products didn't decrease access to the corresponding raw data

# Phase 1 habits of archive data product users



# Phase 1 habits of archive data product users

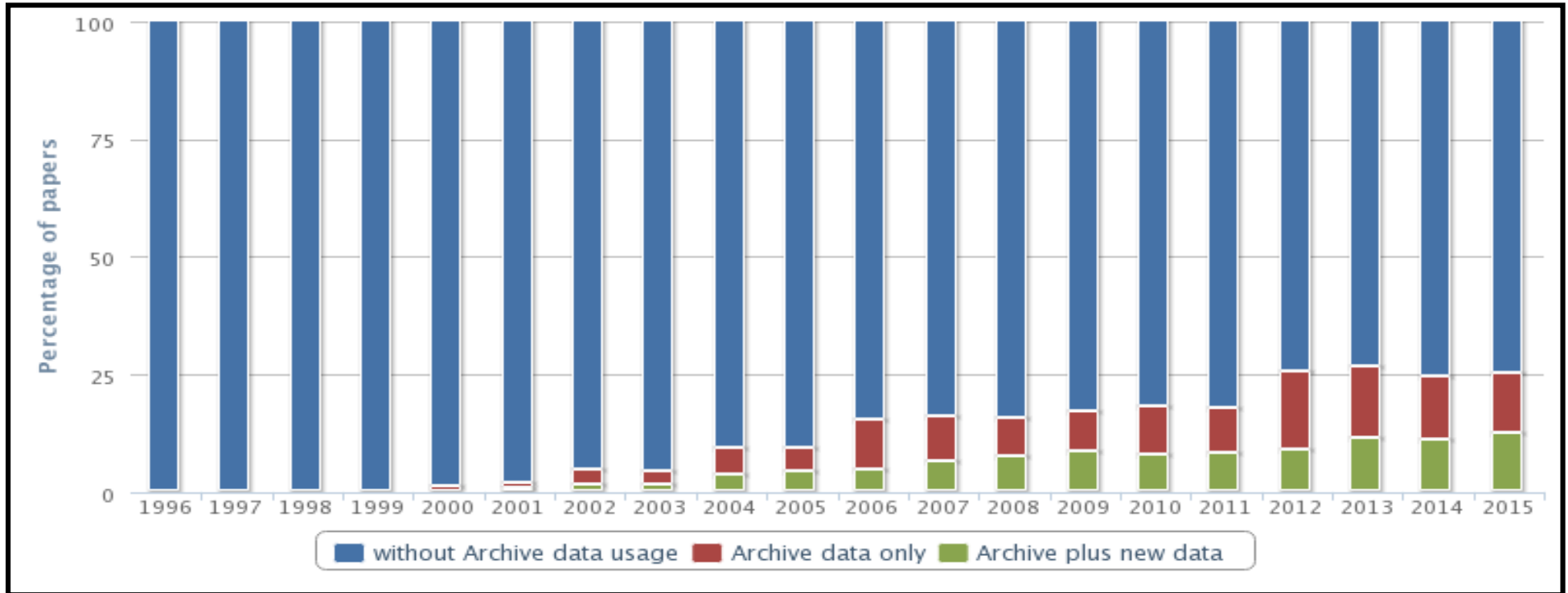


- Phase 3 users: 28% (260) have never applied for time, neither as PIs nor co-Is; 37% (344) have never gotten time, neither as PIs nor co-Is

➤ For comparison, 1/3 of the Phase 1 PIs have never got time

# Archive publications

Source: <http://telbib.eso.org>



- Archive publication: a paper in which none of the authors was part of the original proposal
- After a ramp-up, in the past years archive publications have been a solid 25% of the ESO papers

→ Grothkopf, Friday at 11:55





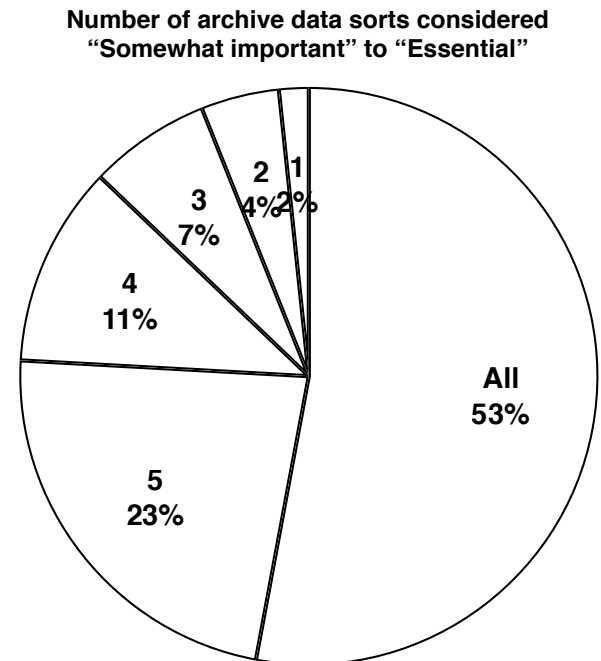
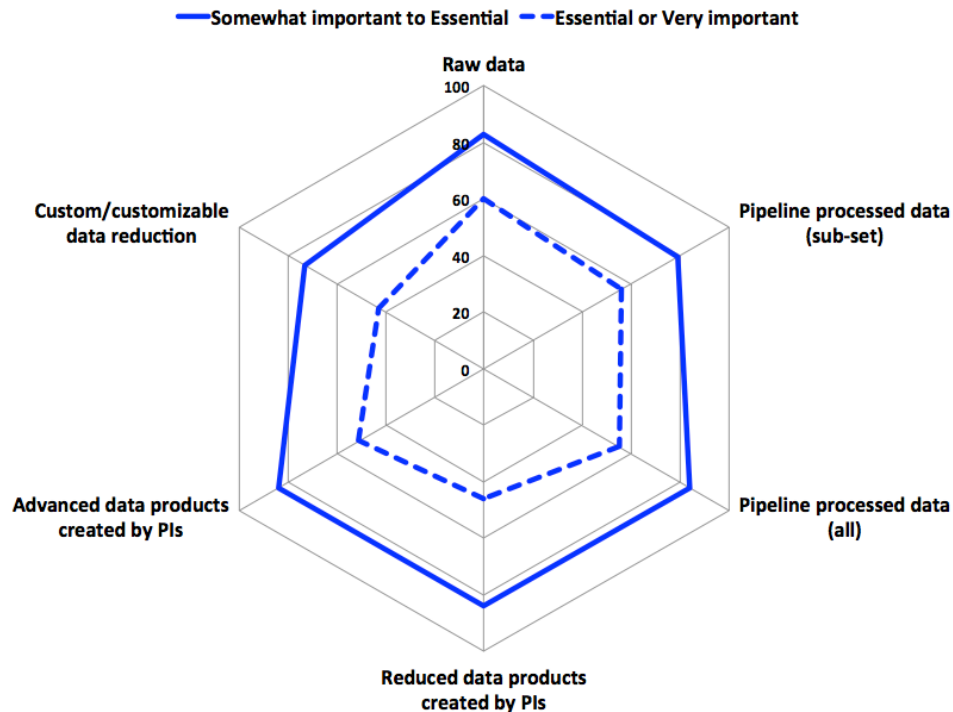
# Finale.

# The future



# Polling the community: ESO2020 questionnaire

- “How important is access to the following sorts of archived data products in order to maximize your scientific productivity?”
  - Statistically (very) significant sample: 1439 answers



■ All different sorts of archived data are deemed important

■ No respondent indicated that archived data is not important



# Archive services: status quo

- **The problem:** as a user, how do I find my way among 25 million files described by 23 billion keywords?
- Data search capabilities of the SAF are currently limited
  - The user-SAF interaction is static and offers limited data discovery capabilities
  - Queries are expressed in terms of technical, rather than physical properties
    - E.g. Technical names of optical components in the light path vs. wavelength coverage and resolution
- This limits the science potential of the SAF itself
  - Especially as its content in terms of ready-for-science data products grows more and more



# Archive services: evolution

- **Guiding principle:** establish a dialogue between the archive researchers, who know their science cases, and the SAF, which knows its content
- First release concentrates on data products, including ALMA
  - Most immediate scientific return, metadata intrinsically more homogeneous, thanks to VO-inspired data standard
- Complex queries on Phase 3 science keywords, which describe the physical properties of the data
  - Signal-to-noise ratio, wavelength coverage and resolution, limiting magnitude, ...
- Added-value services: visualization, facets, ...
- Programmatic access and interoperability



# Science Data and Archives Working Group

- Under the aegis of its Director for Science, ESO has recently completed the prioritization of its programme to ensure that it is well positioned in the likely astronomical landscape of the 2020s
  - [http://www.eso.org/public/about-eso/committees/stc/stc-85th/public/STC-551\\_Science\\_Priorities\\_at\\_ESO\\_85th\\_STC\\_Mtg\\_Public.pdf](http://www.eso.org/public/about-eso/committees/stc/stc-85th/public/STC-551_Science_Priorities_at_ESO_85th_STC_Mtg_Public.pdf)
- Following its recommendations, a Working Group on “Science Data and Archives” is established
  - Maria-Rosa Cioni (Potsdam, UC), Sofia Feltzing (Lund,STC), Françoise Genova (Strasbourg), Bob Mann (Edinburgh), Céline Péroux (Marseille), Martino Romaniello(ESO, Chair), Martin Zwaan (ESO)



# Coda.

## Lessons learnt



# Some lessons learnt

- Data management needs to be highly integrated
  - Observing strategy, data quality, extraction of science content and data dissemination



# Some lessons learnt

- Data management needs to be highly integrated
  - Observing strategy, data quality, extraction of science content and data dissemination
- Invest in quality
  - Not necessarily the ultimate one, but it has to be known and characterized
  - Data product validation greatly enhances their use and legacy value
  - Homogenize the data as early in the data flow as possible





# Some lessons learnt

- Data management needs to be highly integrated
  - Observing strategy, data quality, extraction of science content and data dissemination
- Invest in quality
  - Not necessarily the ultimate one, but it has to be known and characterized
  - Data product validation greatly enhances their use and legacy value
  - Homogenize the data as early in the data flow as possible
- Know your community/work with your community
  - Find out what they need
  - Learn from the experts and make the knowledge available to all
  - Standards need to be defined for maintenance, distribution and fruition
    - Work out the standards with the community