The ESO Archive and Reflex Pipeline Workflows
(or, how do I get data and process it too?)

Mark Neeser (ESO)

Outline

Part 1: In Theory

Getting Data from ESO:
- the contents of the ESO Science Archive Facility (SAF)
- how to access the archive for high-level science data or for raw data

Using Data from ESO:
- Reflex workflows and ESO pipelines
- Reflex workflow basics
- installing Reflex
- a summary of the main features of Reflex

Part 2: In Practise

A Practical Demonstration
- retrieving raw data from the ESO archive
- retrieving science products from the ESO archive
- of the HAWK-I Reflex workflow

The ESO Science Archive Facility (SAF):
==> the operational and technical data archive for the La Silla Paranal Observatory
    (serving 3 + 4 + 2 telescopes and 6 + 12 + 2 instruments)

Data Content (both raw and data products):
~ 500 TB of data consisting of 25 million files (~23 billion database rows of searchable header keywords)

Data Inflow:
~12 TB/month

Data Outflow:
~15 TB/month

Archive Contents

RAW data from the La Silla Paranal Observatory:
==> science, calibration, and technical data

Pipeline products:
mainly science products + ancillary data (standards, catalogues, error maps, etc.)

Two primary sources:

1. products generated at ESO by running the instrument data reduction pipelines with the standard calibration plan
    - these are products that are not tailored to a specific science case, and can be processed further
2. products returned by the community through Phase 3
    - tailored to the science cases of the programmes they originated from (Public surveys, Large Programmes, etc.)
    - include highly processed data (tiles, stacks, catalogues)
    - includes valuable historical data sets (e.g. GOODS, zCOSMOS, . . .)
    - well-documented to summarize the characteristics and limitations of the data

Both are homogeneously accessible through the same archive interface to allow queries on the science and not the data origin

Archive Contents

1. Pipeline products:

many products available since the beginning of science operations

Currently available:
UVES, X-Shooter, HARPS, FLAMES-GIRAFFE

Soon to be available:
HAWK-I, VIMOS (imaging), MUSE, KMOS, and PIONEER

Archive Contents

2. Public Survey products:

- Third public data release of the 6 VISTA public surveys: VIDEO, VHS, VVV, VIKING, UltraVISTA, and VMC
- Second public data release of the 3 VST public surveys: KiDS, VPHAS+, and ATLAS
- Second data release from the 2 public spectroscopic surveys: Gaia-ESO and PESSTO
- 2 surveys with VIMOS just beginning: VANDELS and LEGA-C

sky coverage of ESO public surveys (southern hemisphere & full sky)
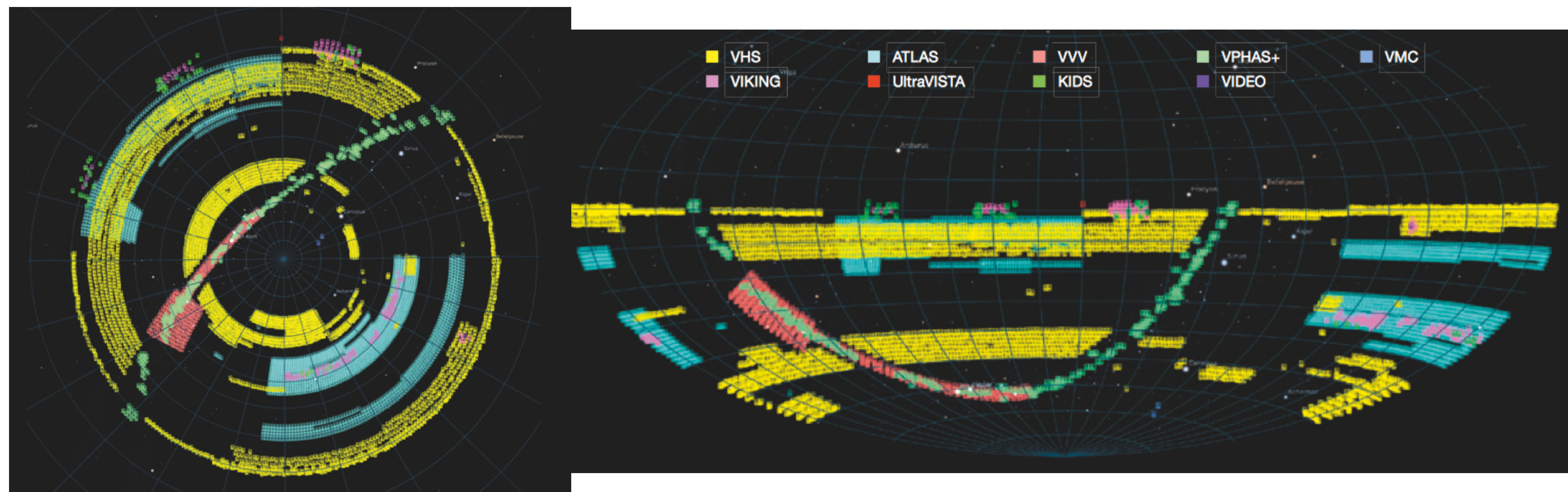
Archive Contents

2. Public Survey products:
e.g. second public data release of the UltraVISTA

from UltraVISTA (http://home.strw.leidenuniv.nl/~ultravista/)
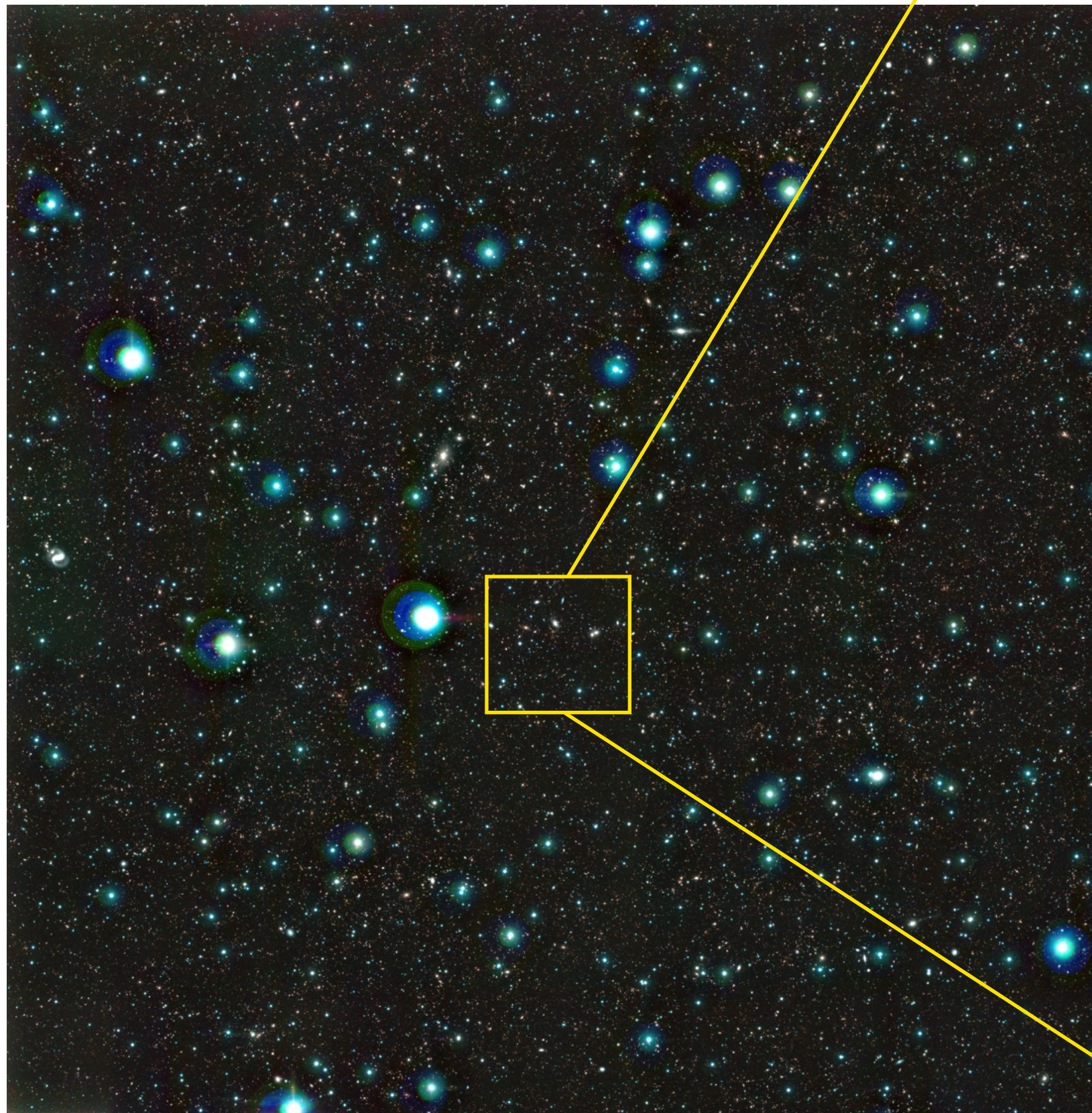
Archive Starting Point

archive.eso.org

1. create your account on the ESO User Portal (not needed for browsing, but for downloads)

2. access RAW data

3. or, access phase 3 data products (images and spectra)

4. or, specific phase 3 catalogues

Welcome to the ESO Science Archive Facility

The ESO Science Archive Facility contains data from ESO telescopes at La Silla Paranal Observatory, including the APEX submillimeter telescope on Llano de Chajnantor. In addition, the raw UKIDSS/WFCAM data obtained at the UK Infrared Telescope facility in Hawaii are available.

The Principal Investigators of successful proposals for time on ESO telescopes have exclusive access to their scientific data for the duration of a proprietary period, normally of one year, after which the data becomes available to the community at large. Please read the ESO Data Access Policy statement for more information, along with the relevant FAQs.

Browsing the archive does not require authentication, but to request and download data you have to log in to the ESO User Portal. Please acknowledge the use of archive data in any publication.

Latest News and Updates

- New Release of PESSTO public survey data (06 Aug 2015)
- First release of the band merged catalogue for the VST Photometric H-alpha Survey of the Southern Galactic Plane (VPHAS+) (30 Jul 2015)
- New Release of Gaia-ESO Spectroscopic Public Survey Data (22 Jul 2015)

More news ...

To browse the archive

Currently, **raw data** and various types of **data products** can be reached via different interfaces:

Category	Access Point	Data collection	Data Type	Instruments
LPO Raw Data	**Raw data** query form (all instruments) **Instrument specific** query forms **Direct retrieval** of raw data by file name	All ESO raw data	Various	Many La Silla Paranal instruments
LPO Data Products [Description of reduced data products types]	**Phase 3 main** query form **Phase 3 imaging** query form **Phase 3 spectral** query form **Phase 3 VIRCAM-specific** query form	Phase 3 Data Products (ESO public surveys; ESO pipeline-reduced products; Large programs: GOODS, zCOSMOS; etc.)	Currently, Imaging and Spectroscopy	Various Pipeline products for UVES, XSHOOTER, HARPS, and more to come.
	Catalogue Facility query interface	Phase 3 Catalogues [ESO User Portal authentication required also when browsing]	Catalogues	Various
	FEROS and HARPS-Polarimetry pipeline processed data query form	FEROS and HARPS-Polarimetry pipeline processed data	Spectroscopy	FEROS, HARPS-Polarimetry, HARPS reduced calibrations (other HARPS see Phase3 above)
	Other Advanced Data Products (available only as downloadable packages, no query form)	Various (30 Doradus, Corot, GaBoDs, etc.)	Spectroscopy Imaging Flux maps	FEROS WFI APEX
	Science Verification, Commissioning, EIS, etc. (no query form)	Full list of available data packages	Various	Many
APEX Quick Look Products	**APEX** query form	APEX	Heterodyne, Bolometer	APEX-2A, LABOCA, SABOCA, SHeFI
LPO Schedule	**Scheduling** query form	ESO Observing Programme Information and Scheduling		All La Silla Paranal instruments, including APEX
ALMA Data	**ALMA Science Archive**	All ALMA data	Cube	ALMA

Access to RAW data

Availability of raw data (Paranal ➙ Garching transfer):

Data transfer Paranal-Garching, June 2014
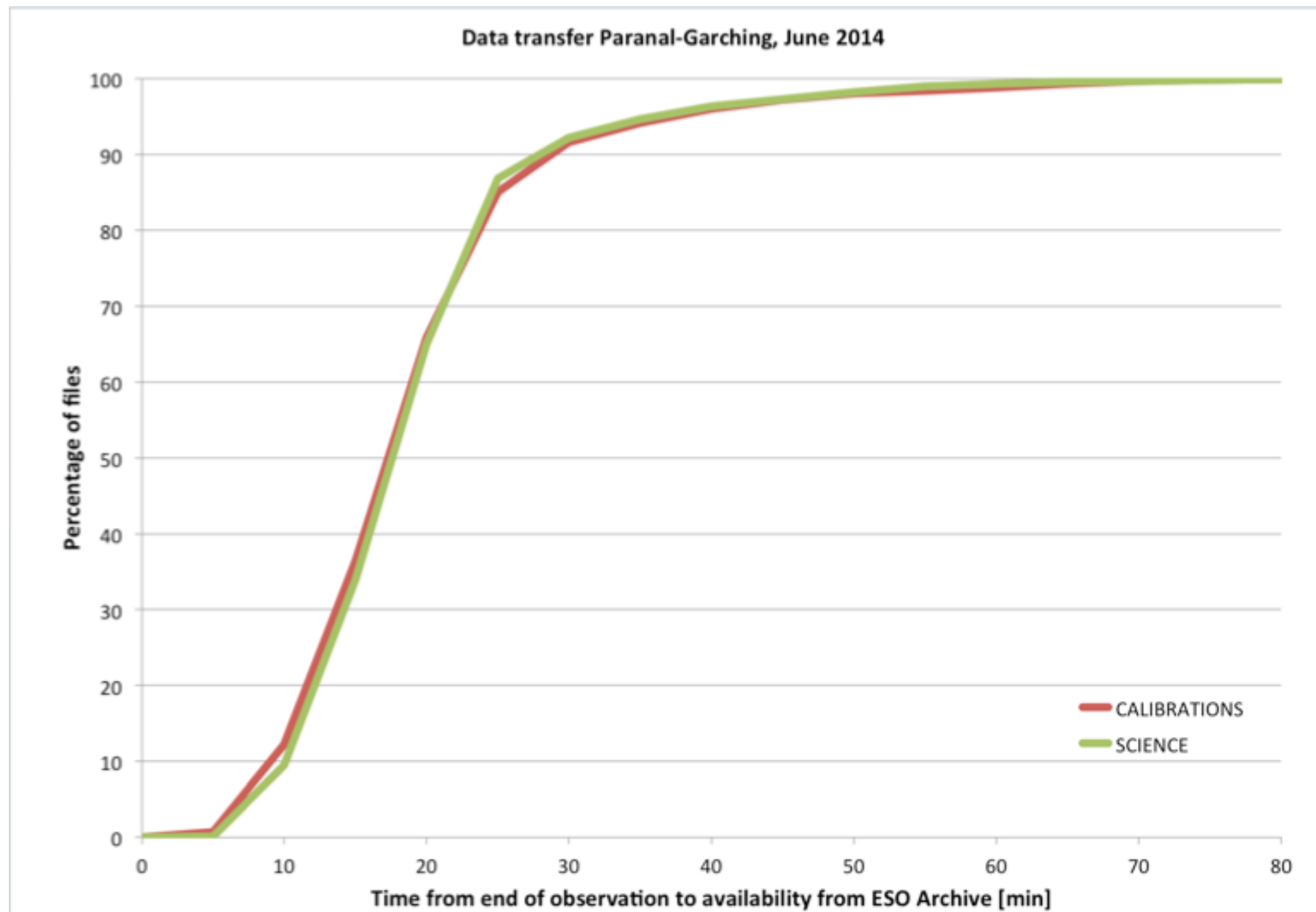
50% within 20^{min}

90% within 30^{min}

98% within 1^{hour}

Archive Content (RAW data)

output preferences

search criteria: target name
RA, DEC
single night
range of nights
programme ID
PI name

instrument specific criteria

Archive Content (RAW data)

search results for:
NGC4038 / FORS2

green indicates publicly
available data
red is proprietary

Archive Content (RAW data)

download searched files:

A new archive feature will match your RAW science frames with:
- associated raw calibrations
- associated master calibrations

For this FORS2 data this means:
- bias

Welcome to the ESO Science Archive Facility

The ESO Science Archive Facility contains data from ESO telescopes at La Silla Paranal Observatory, including the APEX submillimeter telescope on Llano de Chajnantor. In addition, the raw UKIDSS/WFCAM data obtained at the UK Infrared Telescope facility in Hawaii are available.

The Principal Investigators of successful proposals for time on ESO telescopes have exclusive access to their scientific data for the duration of a proprietary period, normally of one year, after which the data becomes available to the community at large. Please read the ESO Data Access Policy statement for more information, along with the relevant FAQs.

Browsing the archive does not require authentication, but to request and download data you have to log in to the ESO User Portal. Please acknowledge the use of archive data in any publication.

Latest News and Updates

- First release of PESSTO spectral data products (20 Jan 2014)
- DSS and Skycat catalogues services migrated (17 Jan 2014)
- New data of the VISTA Deep Extragalactic Observations Survey in the VIDEO-XMM field released (17 Jan 2014)

More news ...

To browse the archive

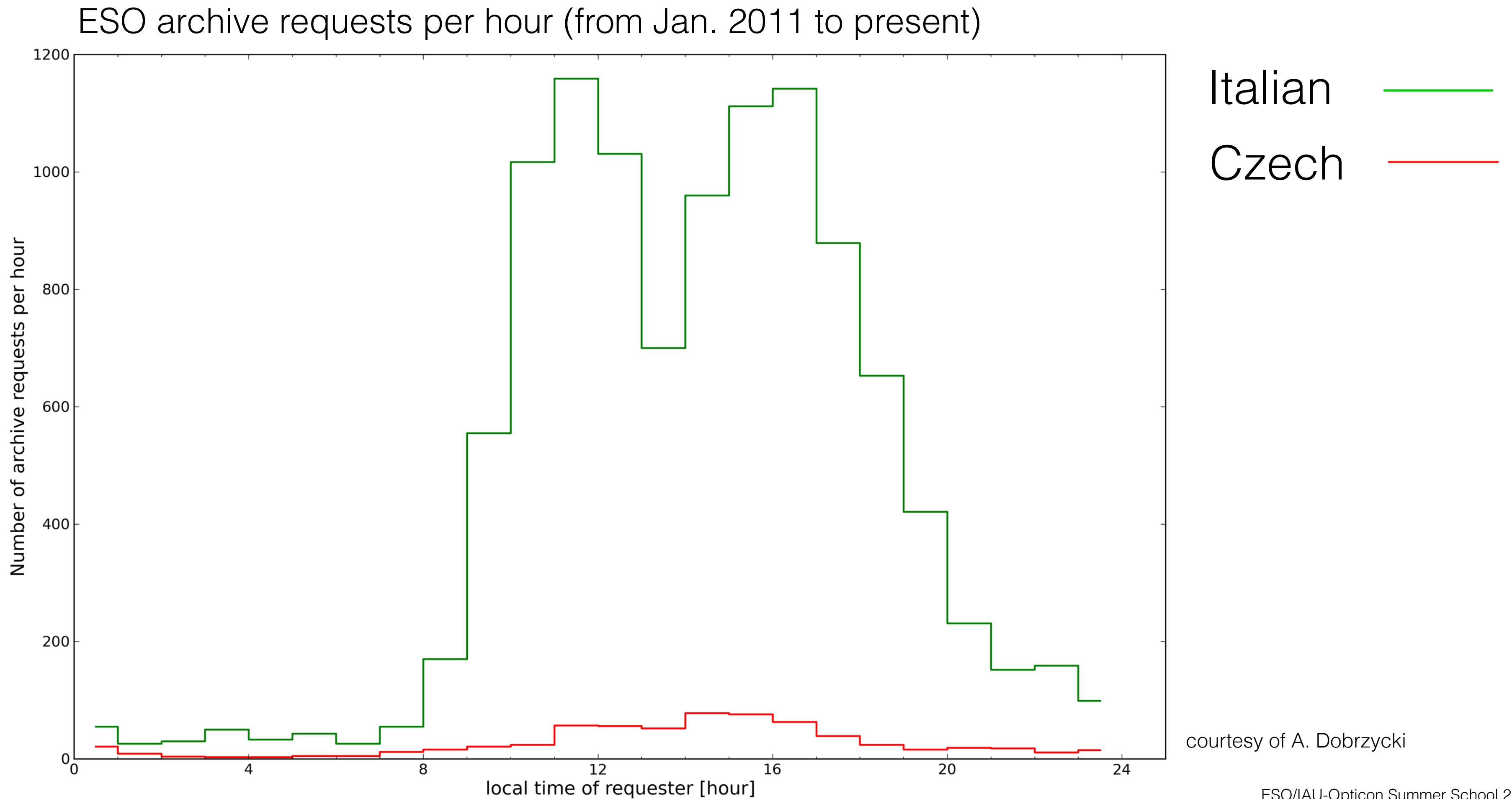Currently, **raw data** and various types of **data products** can be reached via different interfaces:

Category	Query Forms	Data collection	Data Type	Instruments
LPO Raw Data	**Raw data** query form (all instruments) **Instrument specific** query forms **Direct** retrieval of raw data by file name	All ESO raw data	Various	Many La Silla Paranal instruments
LPO Data Products	**Phase 3 main** query form **Phase 3 imaging** query form **Phase 3 spectral** query form **Phase 3 VIRCAM-specific** query form	Phase 3 Data Products (ESO public surveys, large programs, pipeline products, etc.)	Currently, Imaging and Spectroscopy	Currently, VISTA/VIRCAM, VST/OmegaCAM, UVES pipeline products, etc.
	Catalogue Facility query interface	Phase 3 Catalogues [ESO User Portal authentication required also when browsing]	Catalogues	Currently, VISTA/VIRCAM, FEROS
	Advanced Data Products query form	GOODS (C.Cesarsky)	Imaging, Spectroscopy	FORS2/ISAAC/VIMOS
		zCOSMOS (S.Lilly)	Spectroscopy	VIMOS
		Observation of Corot astroseismologically-selected HD stars (E.Poretti)	Spectroscopy (time series)	FEROS
		Time-domain survey of NGC 2547 (S.Aigrain)	Imaging	WFI
	FEROS/HARPS pipeline processed data query form	FEROS/HARPS pipeline processed data	Spectroscopy	FEROS, HARPS
	Science Verification, Commissioning, EIS, etc.	Full list of available data packages	Various	Many
APEX Quick Look Products	**APEX** query form	APEX	Heterodyne, Bolometer	APEX-2A, LABOCA, SABOCA, SHeFI
LPO Schedule	**Scheduling** query form	ESO Observing Programme Information and Scheduling		All La Silla Paranal instruments, including APEX
ALMA Data	**⮕ ALMA Science Archive**	All ALMA data	Cube	ALMA

The ESO Science Archive Facility was developed in partnership with the *Space Telescope – European Coordinating Facility (ST-ECF)*. It was operated jointly until the ⮕ closure of the ST-ECF in December 2010.

courtesy of M. Romaniello

archive.eso.org

Archive Usage (an aside)

ESO archive requests per hour (from Jan. 2011 to present)
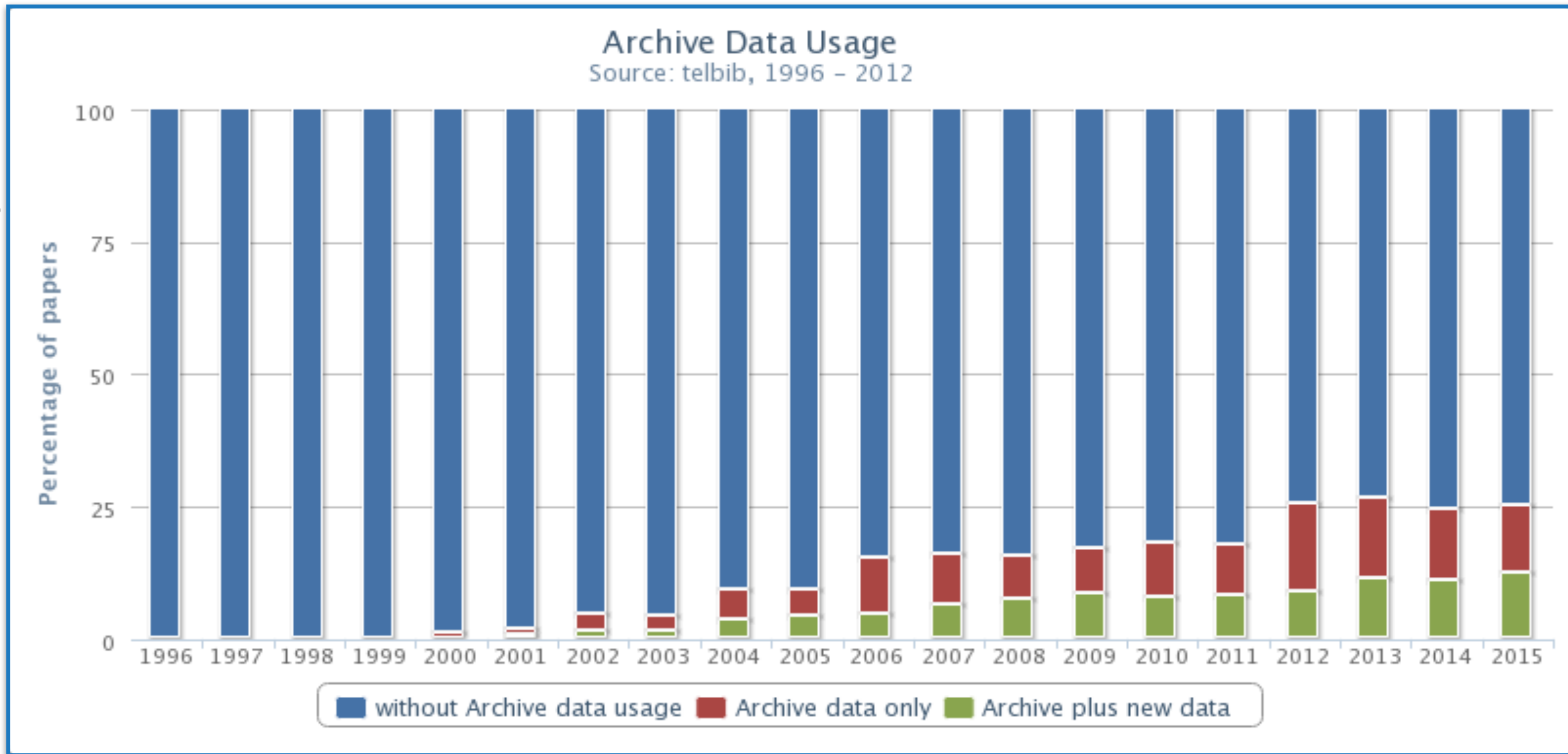

Italian ——
Czech ——

courtesy of A. Dobrzycki

Conclusions: Archive Value

Archival (re)use of data is/will be a major mode for modern astronomy:

- gives anyone the ability to scrutinize any claimed result ==> the foundation of the scientific method

- data recycling increases its initial value

- allows genuinely new science (e.g. statistically meaningful samples from scattered, multiple programmes, or time-domain studies.

- allows for fast-track science (submitting observing proposals and getting data takes a long time)

- lowers the access threshold to data (==> a new community of users)

- allows a PI to assess the feasibility of new observations (preparation for phase 1 and phase2).

Conclusions: Archive Value

Archive Data Usage
Source: telbib, 1996 – 2012

Legend: without Archive data usage — Archive data only — Archive plus new data

An archive publication is defined as a paper in which none of the authors was part of the original proposal
==> a steady 25% of ESO refereed publications since 2012
==> equivalent to one VLT 8.2 m telescope!

Processing Your Data

By 2016 ESO/Paranal will be offering ~20 instruments to the community

http://www.eso.org/public/teles-instr/vlt/vlt-instr/
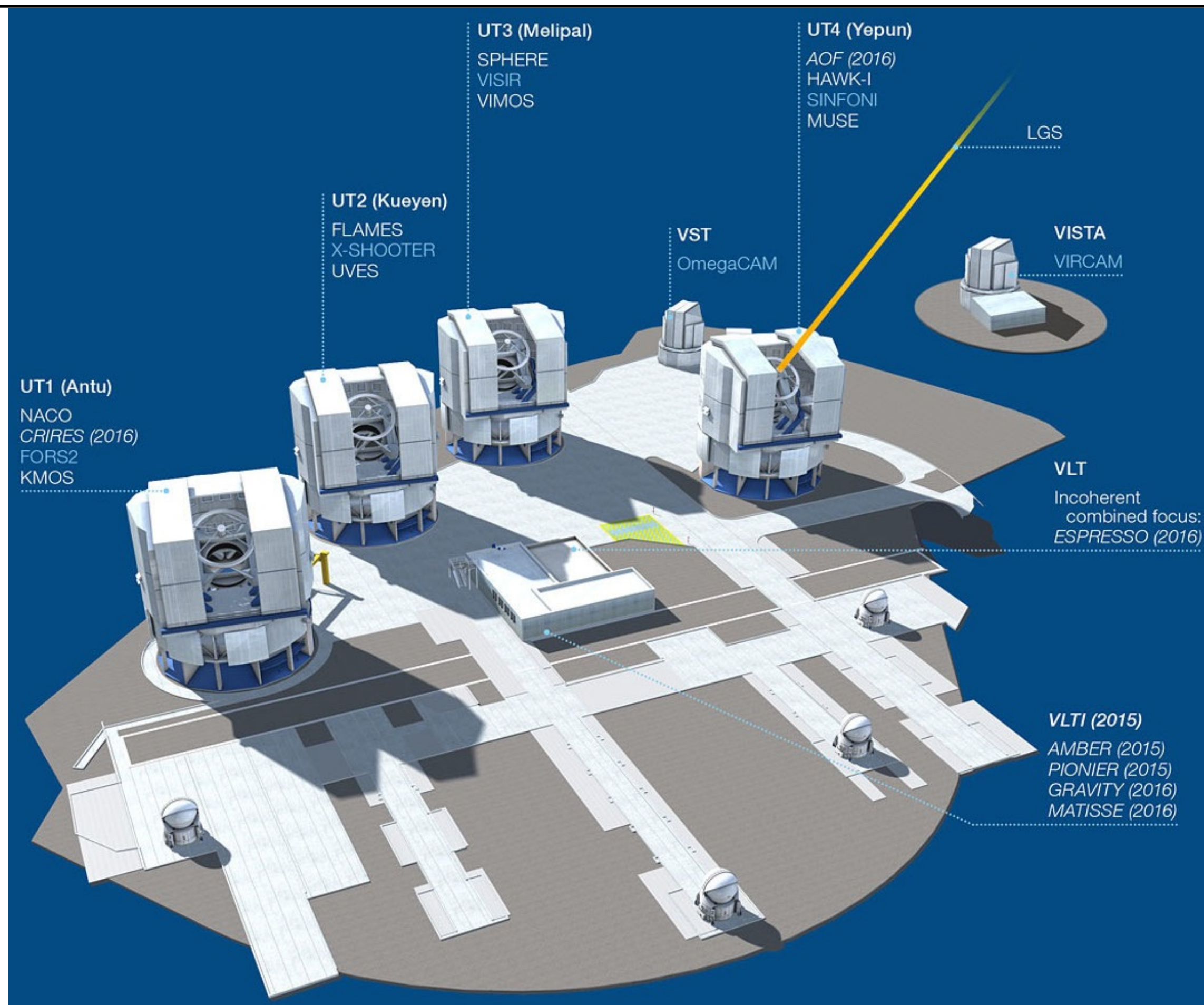
Processing Your Data

These instruments are becoming very complex

==> complex data formats (e.g. multi-layered MEF files)
==> complex data dependencies (science <—> calibrations)
==> complex pipelines

Traditionally, the way to process your raw data involves:

- a steep learning curve to become an expert in the instrument (read a lot of users'/instrument/pipeline manuals).
- write your own scripts to sort, select, and process the data (IDL, IRAF, python, C++, etc.).
- ~trial-and-error your way to some reasonable results.

There is now a better way . . .
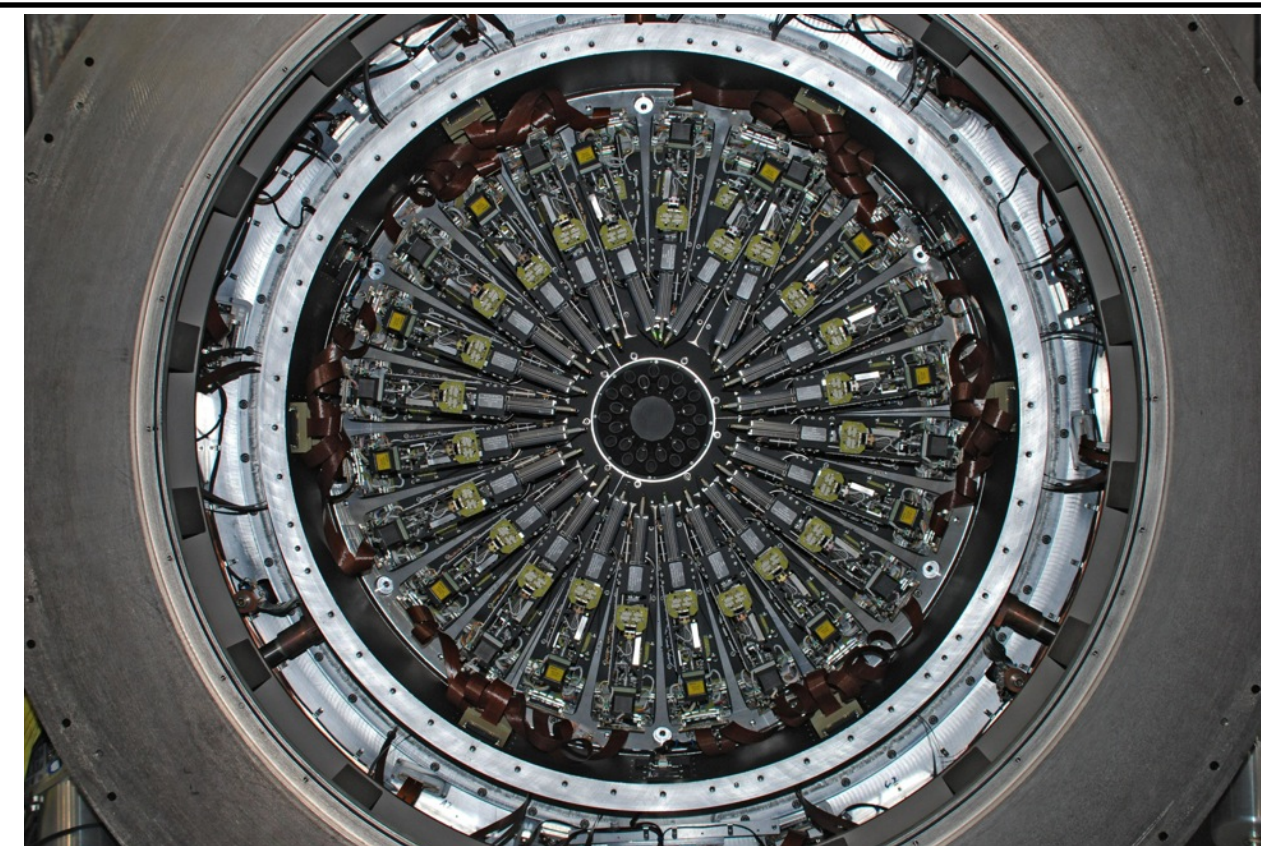
e.g. KMOS: 24 fibre arms feeding 3 spectrographs
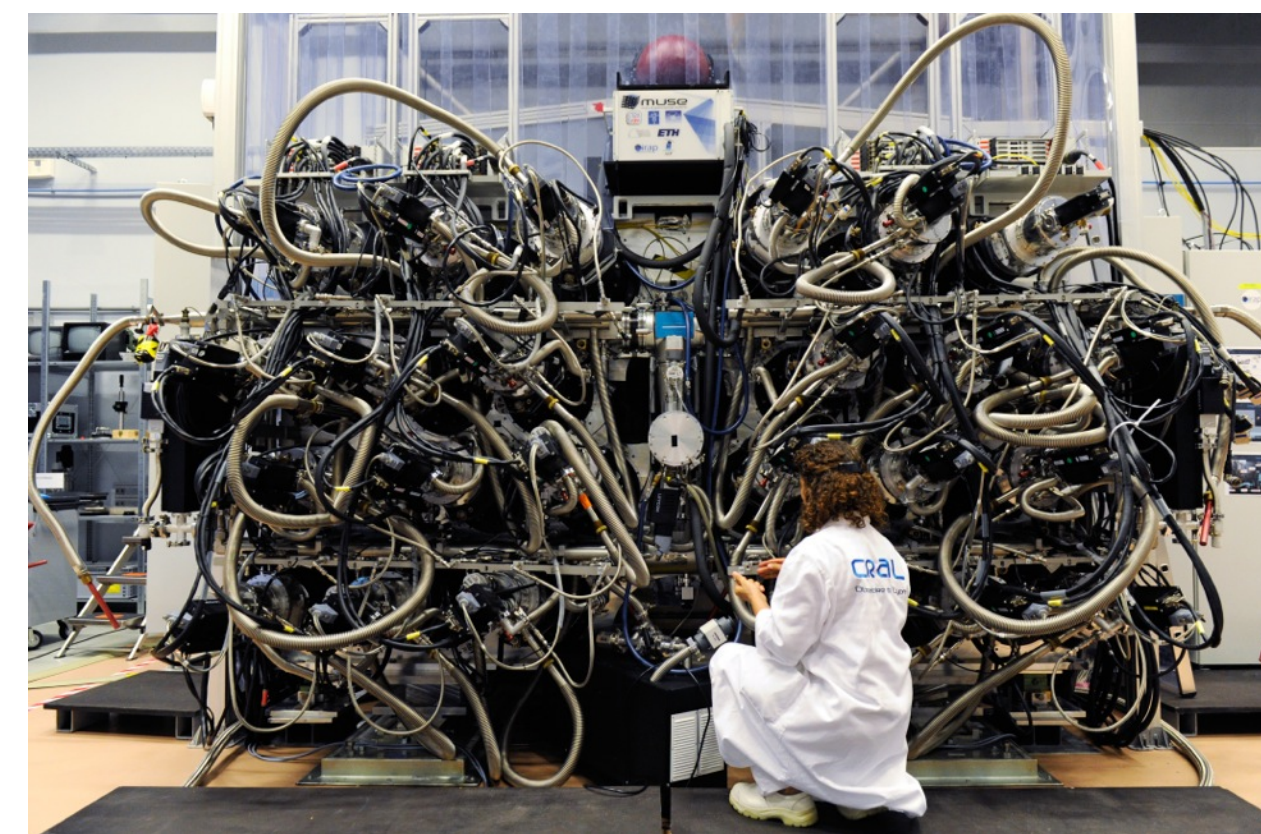
e.g. MUSE: an integral field spectrograph with 24 IFU's each of which is sliced into 48 mini slits

Processing Your Data Using ESO Reflex

ESO Reflex makes pipeline recipes accessible to the general user

- each Reflex workflow has been designed and tested by instrument & pipeline 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)
- visual understanding

Design Goals

- will organize the data for you
- pipeline will run with a single *click*
- you can monitor the progress of the workflow
- will do the book-keeping for you
- will allow user interaction and modification of pipeline parameters
- allows insertion of User procedures in any language

KMOS Reflex Workflow

ESO Reflex is now the recommended environment to run VLT pipeline recipes

ESO Reflex
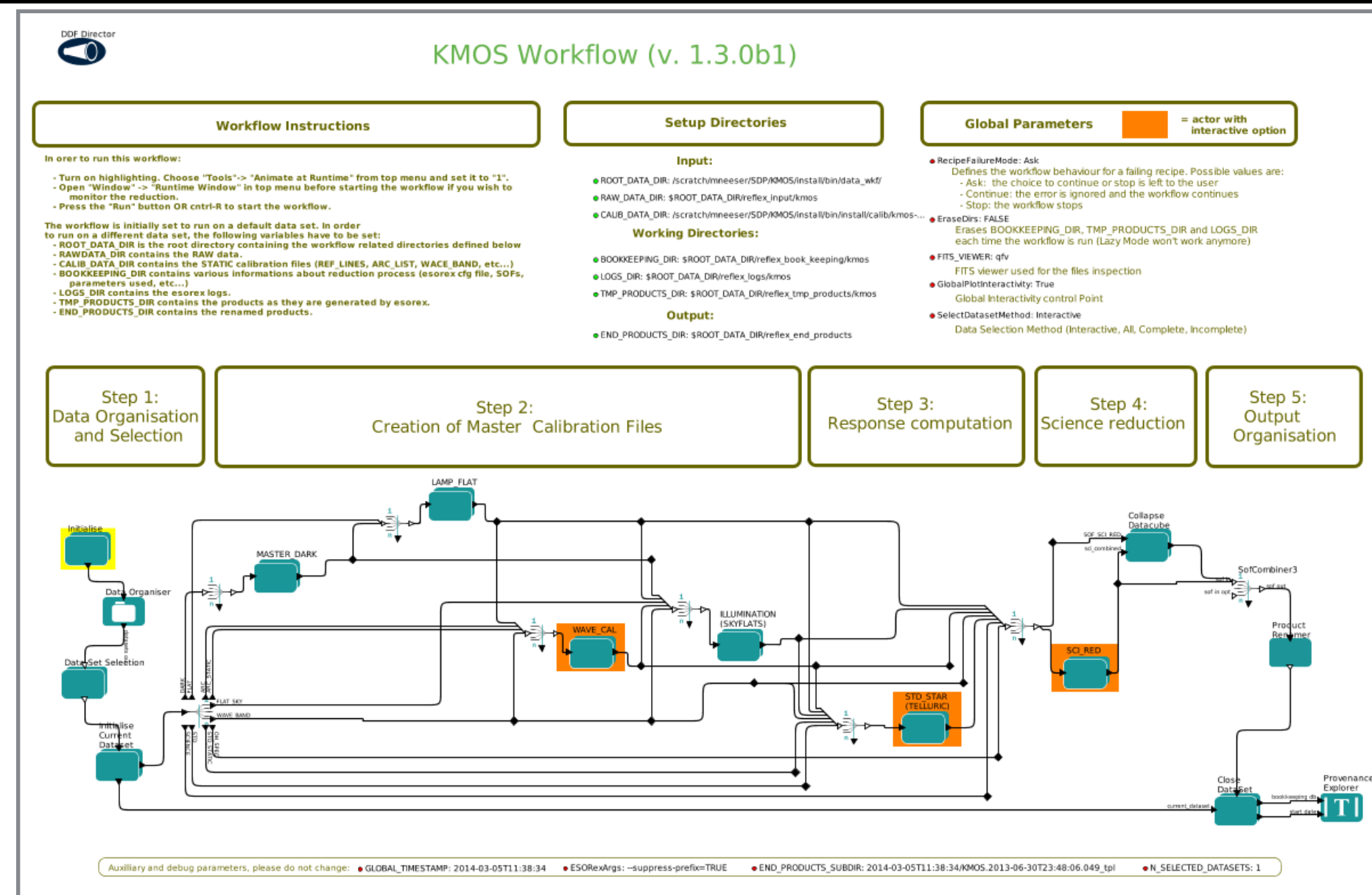
Reflex is based on the Kepler Workflow engine

- https://kepler-project.org

- Kepler is freely available under the BSD license.

- used in the life sciences, engineering, climate sciences, ecology, geology, and now in astronomy.

- provides a graphical user interface (Java) to create workflows.
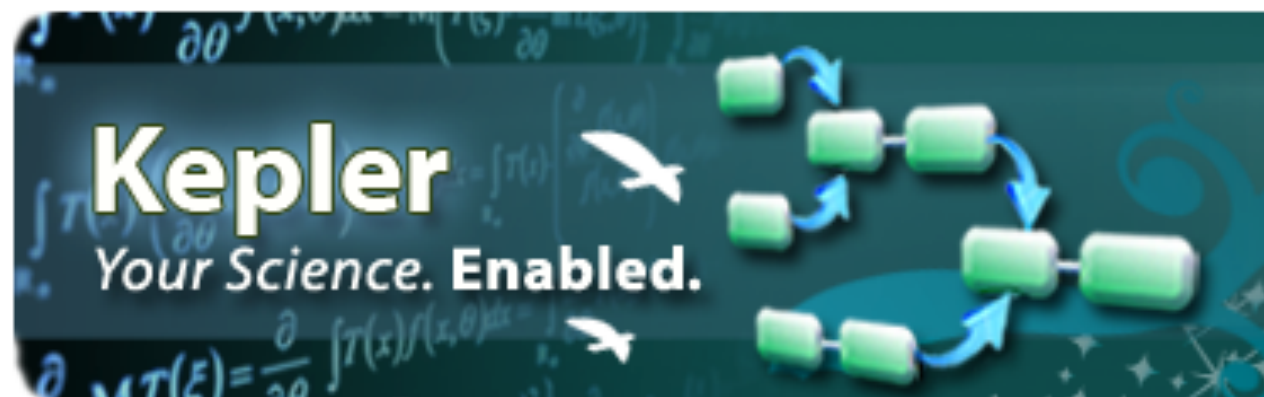
ESO Reflex

List of current and future ESO Reflex Workflows:

- FORS (imaging, spectroscopy, and MOS)
- KMOS
- MUSE
- SINFONI
- UVES
- VIMOS (MOS)
- XSHOOTER

- HAWK-I
- VIMOS (imaging)   ==> by the end of 2015
- VIRCAM
- FLAMES-GIRAFFE
- VISIR

- ESPRESSO
- SPHERE
- GRAVITY
- MATISSE

All **new** ESO instruments are required to be delivered with a pipeline & Reflex workflow

ESO Reflex

Documentation and links:

Reflex start-up page: http://www.eso.org/sci/software/reflex/
Reflex paper: http://www.eso.org/sci/software/pipelines/documents/reflex_aa.pdf
available workflows & pipeline manuals: http://www.eso.org/sci/software/pipelines/

Installation:

start here: http://www.eso.org/sci/software/pipelines/#installation_instructions

for Linux:
Fedora 20/21/22:   — installation rpm's
all others: — download *install_esoreflex* script and execute on the command line;
                     follow the instructions to download and install the workflow of your choice

Mac OS 10.9+:   — install using *macports*
                   — http://www.eso.org/sci/software/pipelines/installation/macports.html

ESO Reflex: getting started

each Reflex workflow installation includes:

- the instrument pipeline
- a demonstration data set
- the Reflex workflow (as an .xml file)
- a tutorial manual

initialise via the command line:

./esoreflex &

- - begins with an empty viewer.
- - select <instrument>.xml to initialise the workflow.
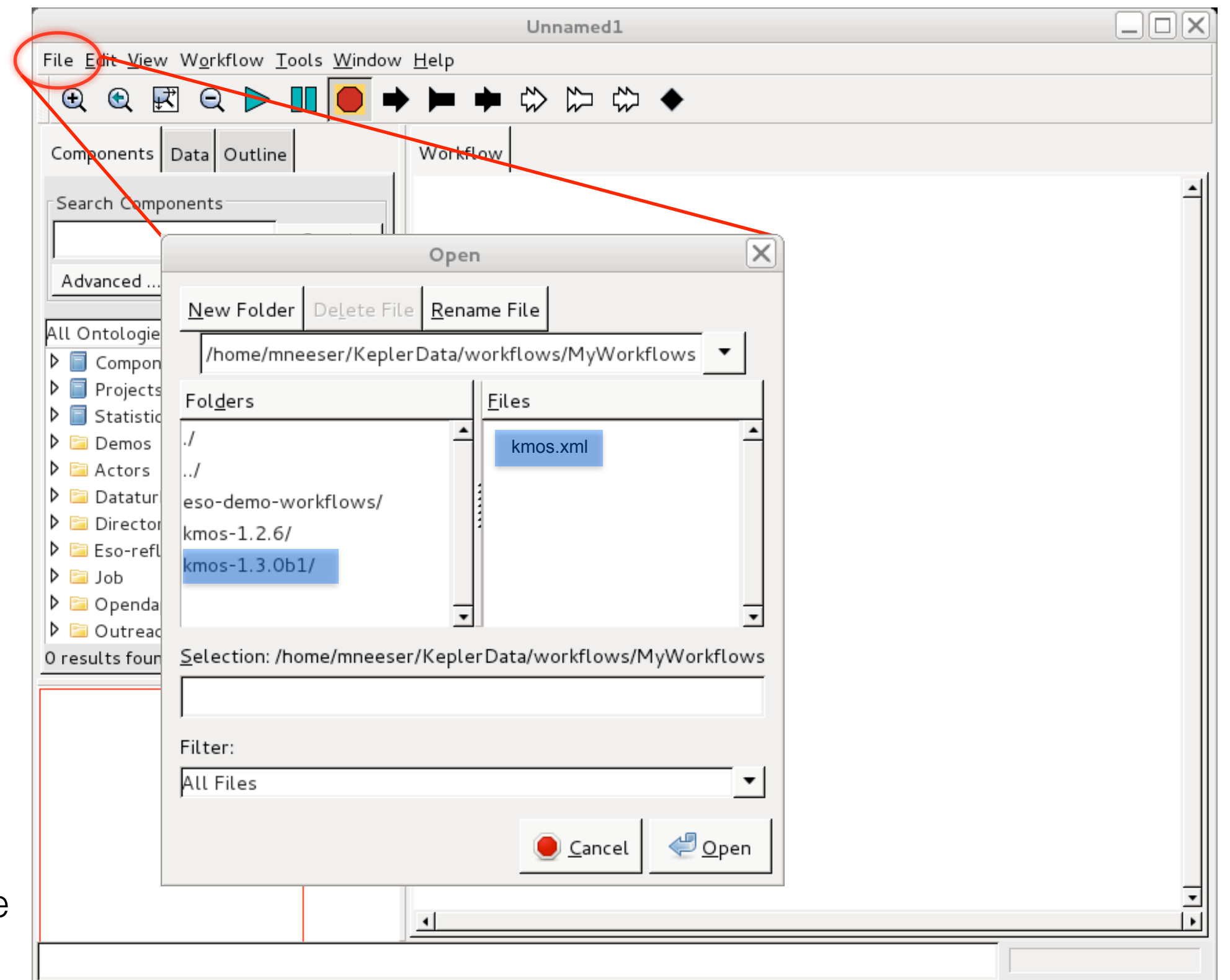
ESO Reflex: a summary of the main features
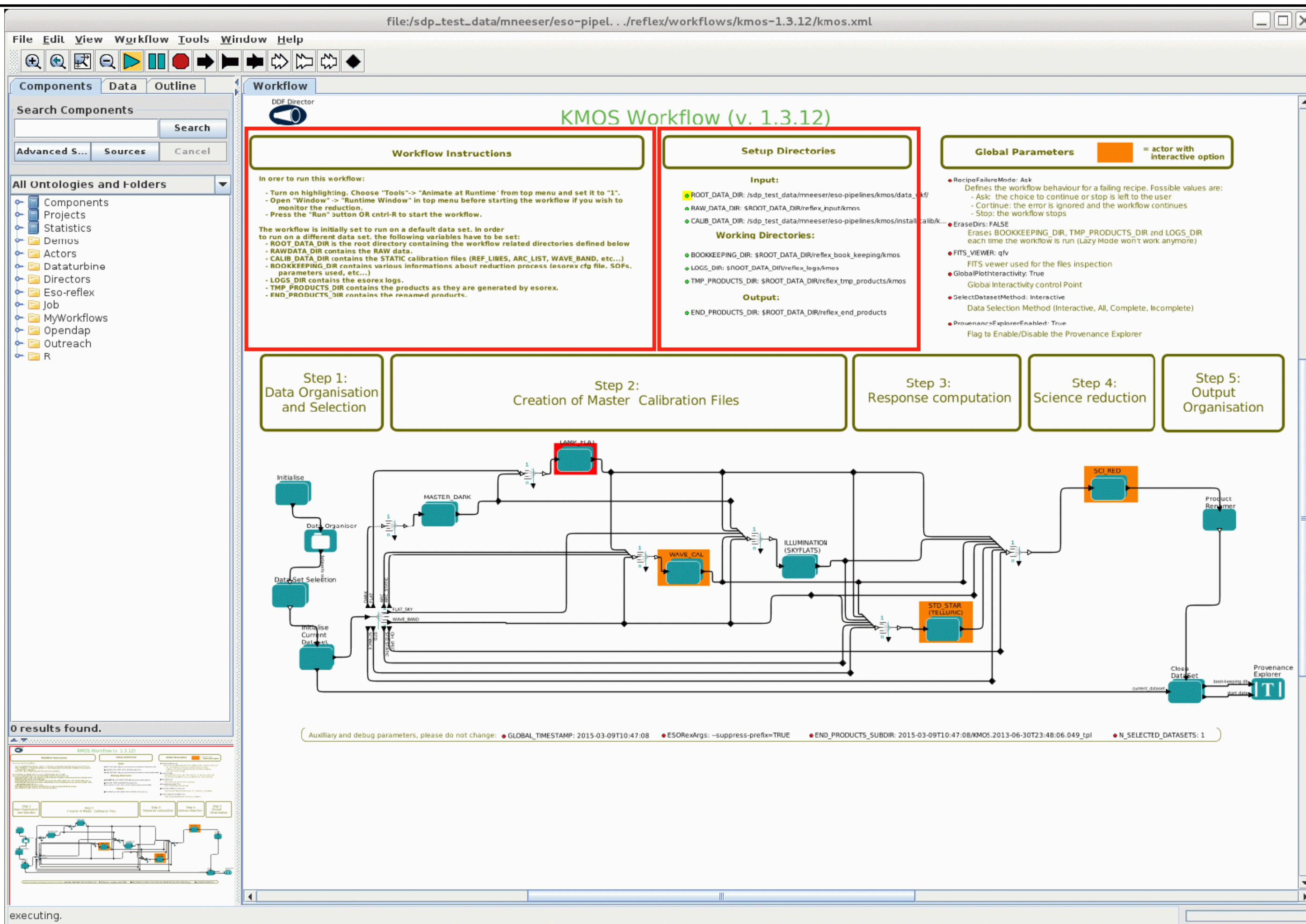
e.g. the KMOS Reflex Workflow

- to run the demonstration data set just click start: ▶

- to run the workflow on your own data follow the instructions to change the setup directories
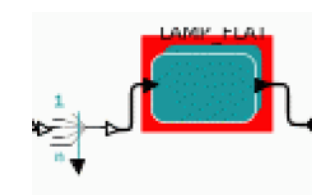
high-lighted Actor shows where in the workflow you are
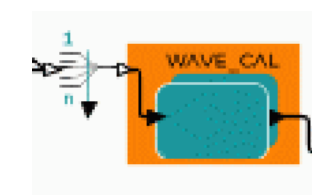
indicates an Actor with an interactive option

ESO Reflex: a summary of the main features

- **DataOrganiser**

  - organises all input data (science & calibrations) into groups (datasets)
    that can be processed independently by the workflow.
    Each dataset has a tree structure and represents the complete calibration
    cascade.
    The datasets are organised, classified, and associated using the data headers

  ==> any science frame will be optimally matched to its required
       calibration frames (e.g. for KMOS it will:
       DIT-matched darks, arc lamps, flats, ArNe
       line lists, illumination correction flats, telluric standard stars,
       atmospheric models, spectral type look-up tables, etc.)
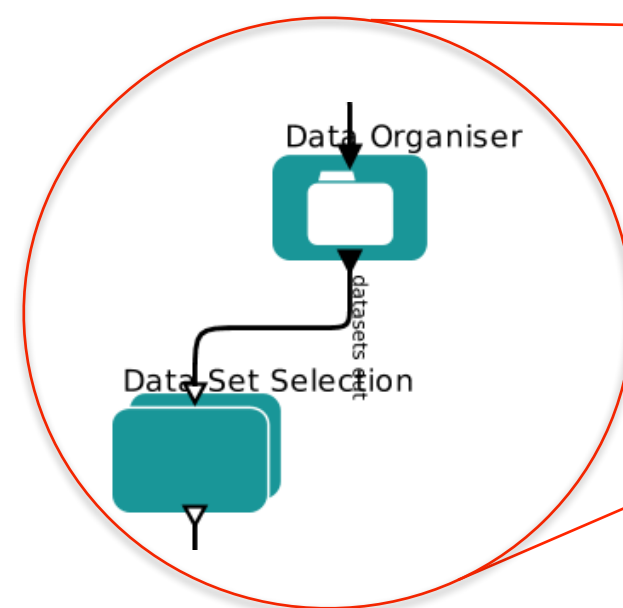
- **DataSetSelection**

  - will list all unique data sets found by the DataOrganiser and allow the
    user to select which ones to process.
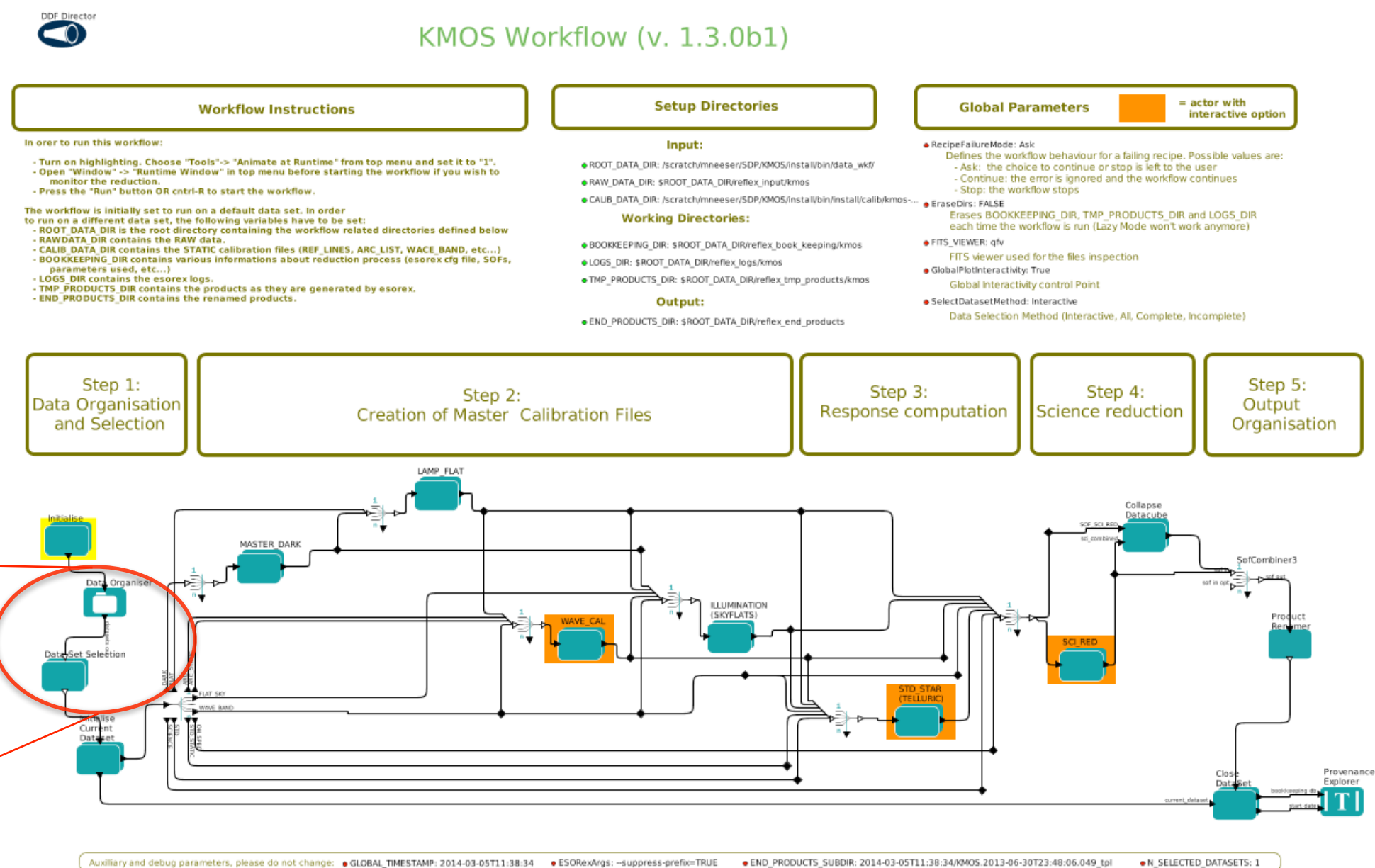
ESO Reflex: a summary of the main features

then, the pipeline processing begins:

- **Master Dark**
  - the DIT-matched raw darks are processed into master dark frames and hot pixel masks.

- **Lamp Flat**
  - the filter-matched lamp flats are processed into master flat frames and cold pixel masks.

- **Wavelength Calibration**
  - a wavelength solution is found for the filter-matched arc frames, which are then processed into reconstructed arc line frames and a wavelength solution look-up frame.
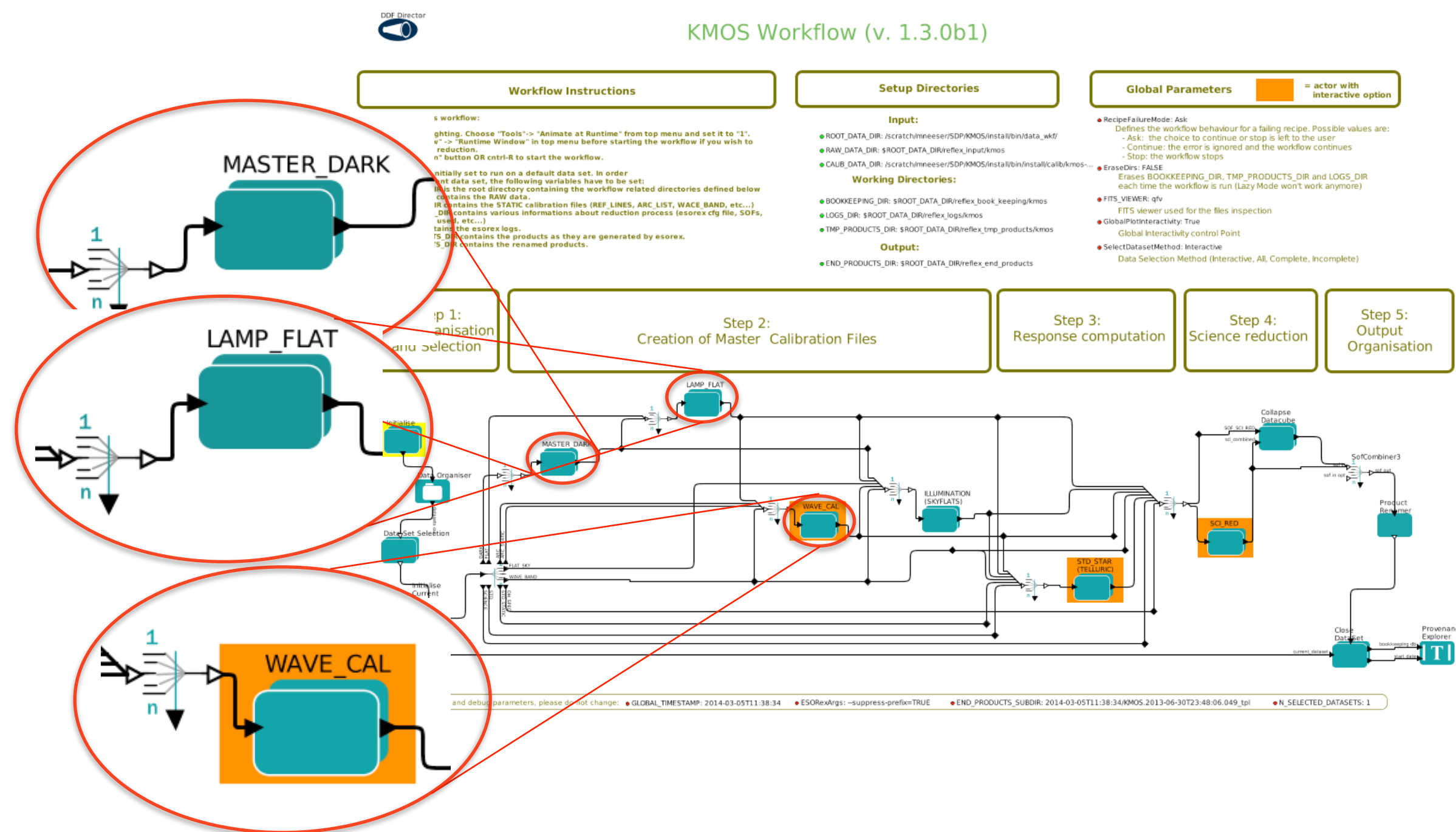
ESO Reflex: a summary of the main features

Interactive Actors

- ### Wavelength Calibration

  - reveals reconstructed arc lamp frames for each detector and can be selected for each of the six calibration rotator angles (0°, 60°, 120°, 180°, 240°, and 300°)

  mouse-over reveals pipeline QC keyword values

  ```
  ESO QC ARC AR POS MEAN:   -0.620317
  ESO QC ARC AR FWHM MEAN:   1.867742
  ESO QC ARC NE POS MEAN:    0.313877
  ESO QC ARC NE FWHM MEAN:   1.823414
  ```

Modifiable pipeline parameters

can be changed and re-run in pipeline to optimise results.
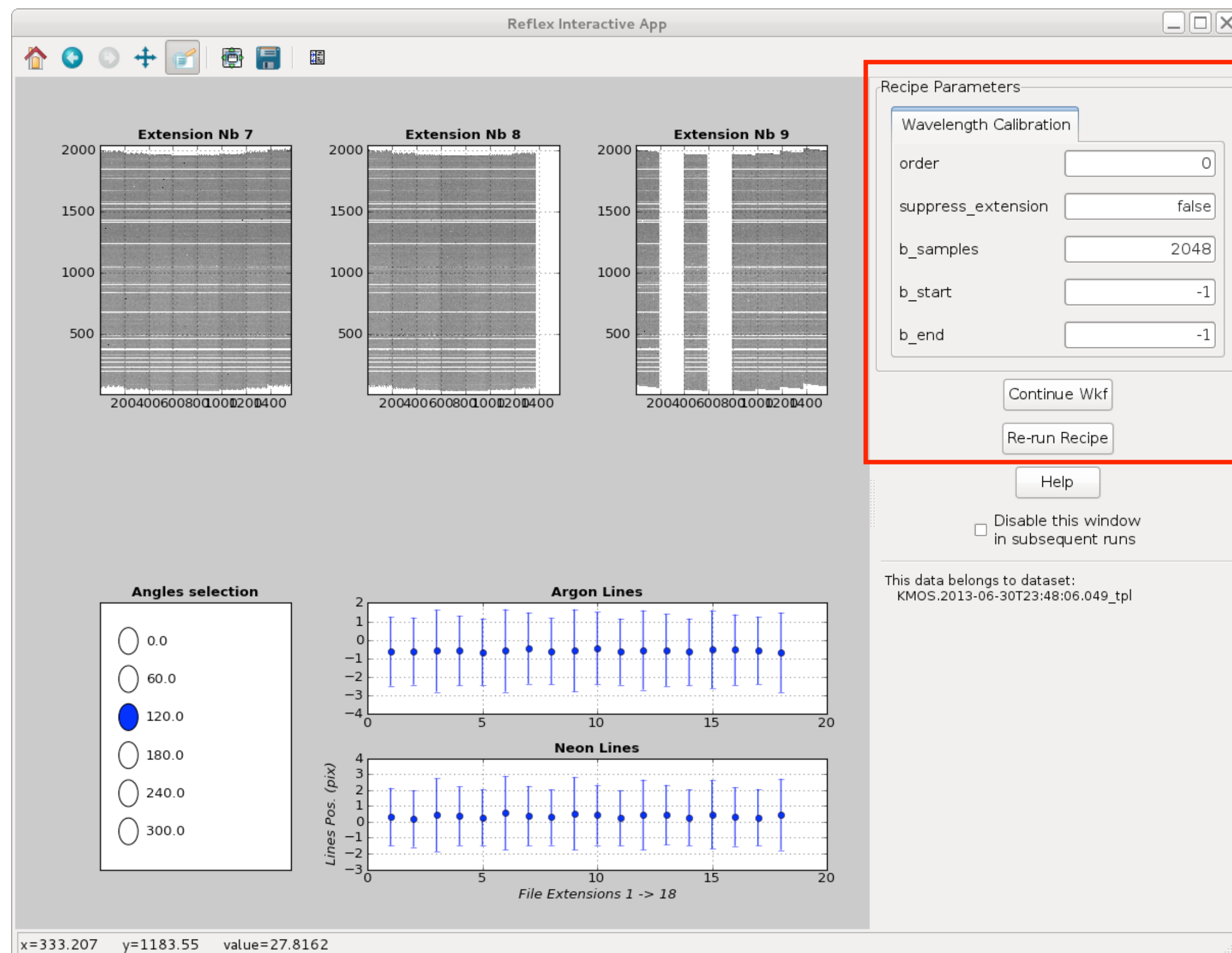
ESO Reflex: a summary of the main features

the processing continues:

- **Illumination Correction**
  - filter-matched sky flats are used to compute an illumination correction for each IFU.

- **Telluric Standard Star**
  - filter-matched standard star fields processed into a median cube-collapsed image and an extracted spectrum.
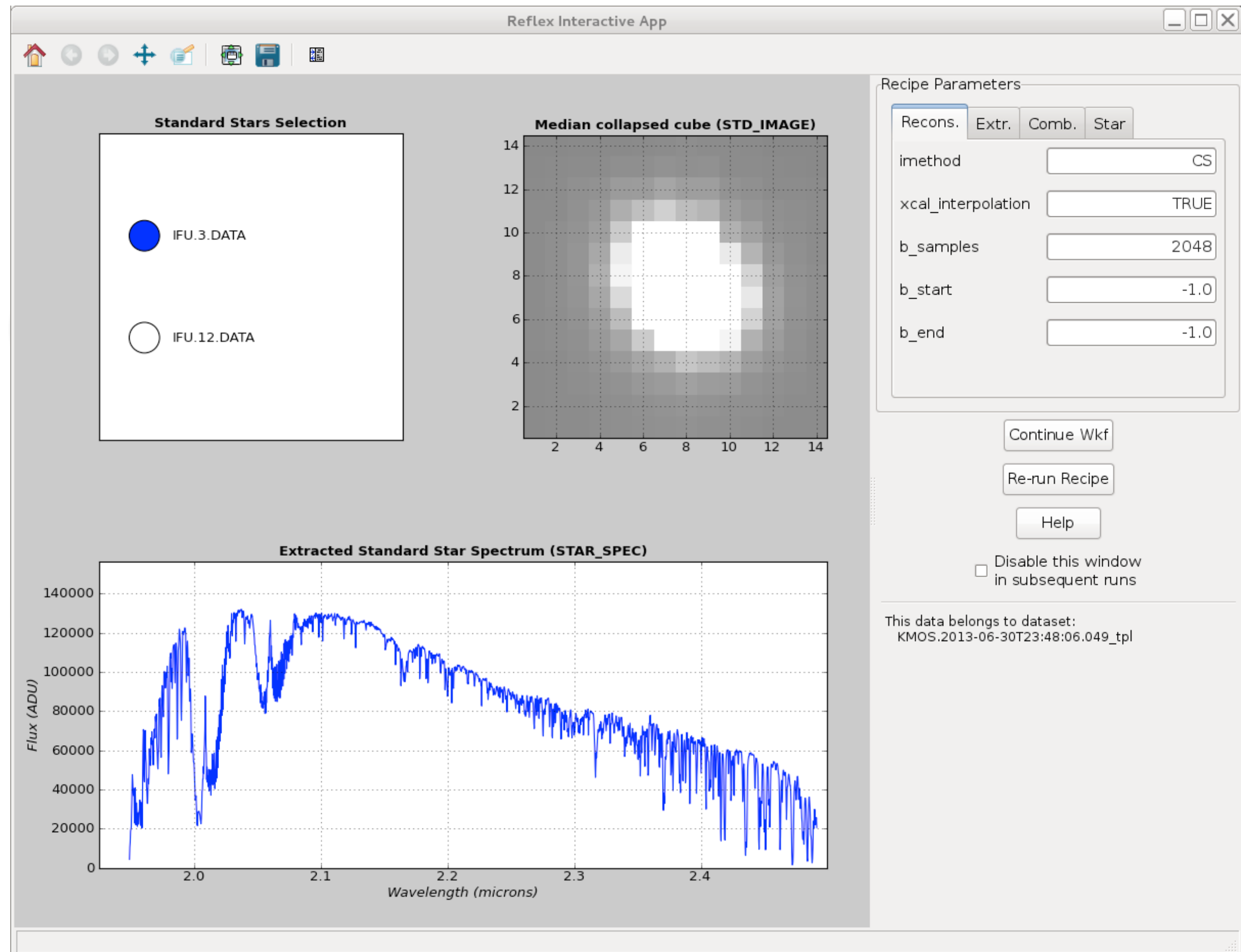
ESO Reflex: a summary of the main features

the processing continues:

- **Science Reduction**
  - science frames are flat-fielded, wavelength calibrated, illumination corrected (optional), flux-calibrated (optional), reconstructed and combined.
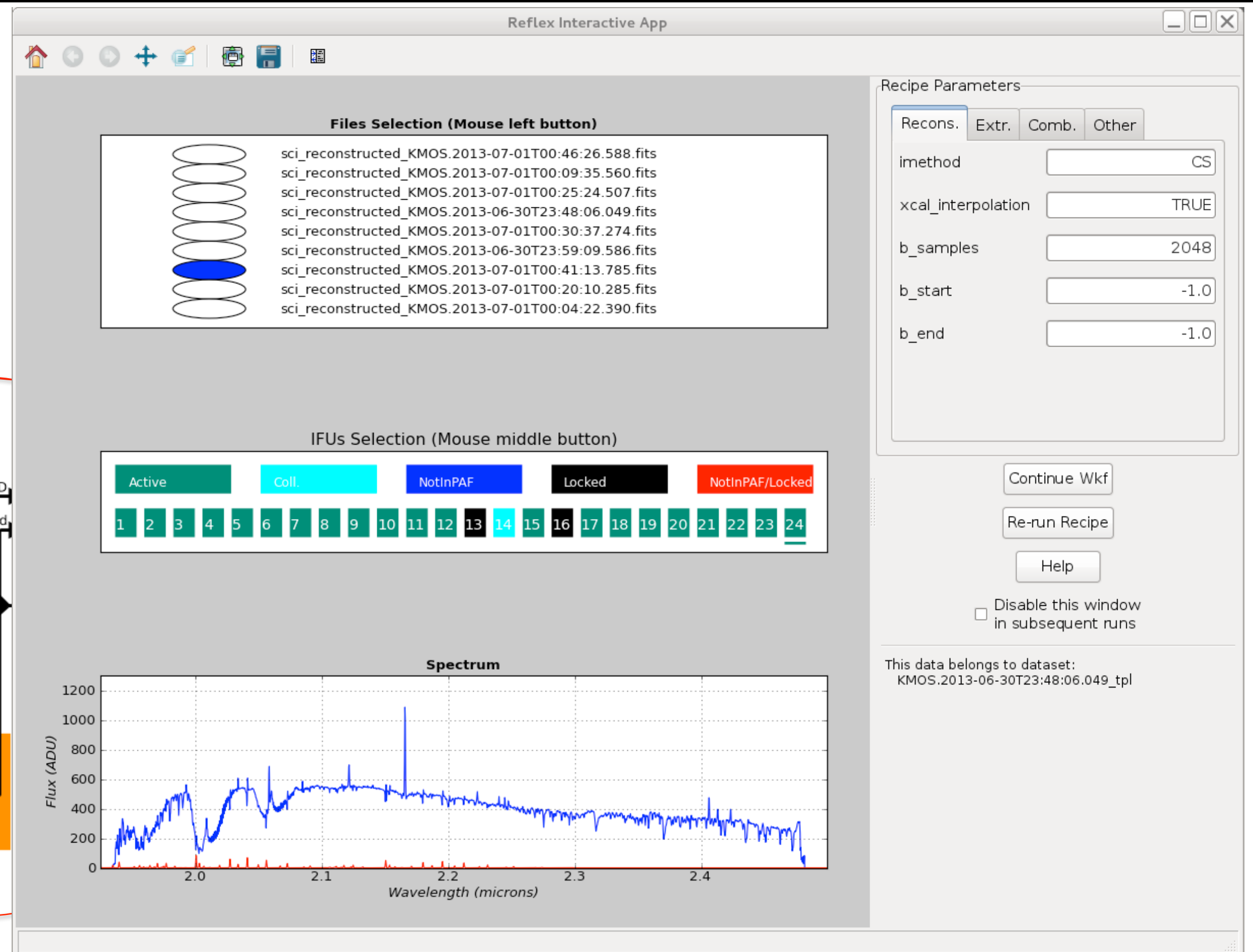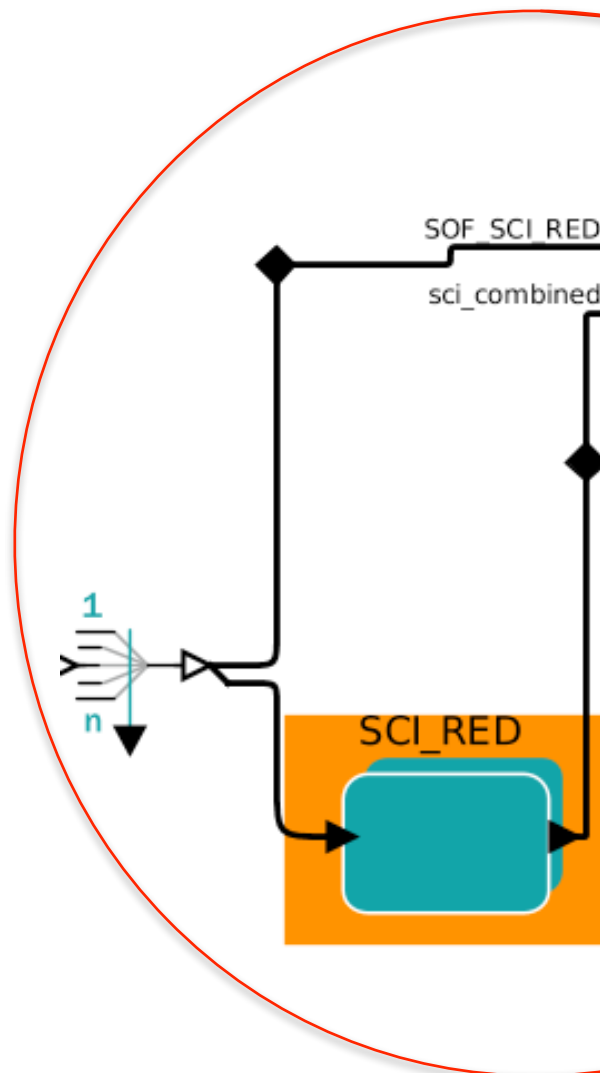
ESO Reflex: a summary of the main features

• **Product Provenance**

- the Reflex workflow finishes with a product provenance window in which all pipeline products can be tracked to the raw & calibration frames that were used in their creation.

ESO Reflex: a summary of the main features

Results and Products:

- all pipeline products, book-keeping, and pipeline logs are saved to the directories specified in the setup directories.

- each run through the workflow with differing inputs will be saved in a directory with a different time-stamp.

- Lazy Mode (**on**: calibrations from previous runs are re-used; **off**: calibrations are recreated).

- interactive actors can be turned off for batch-mode execution of the pipeline.
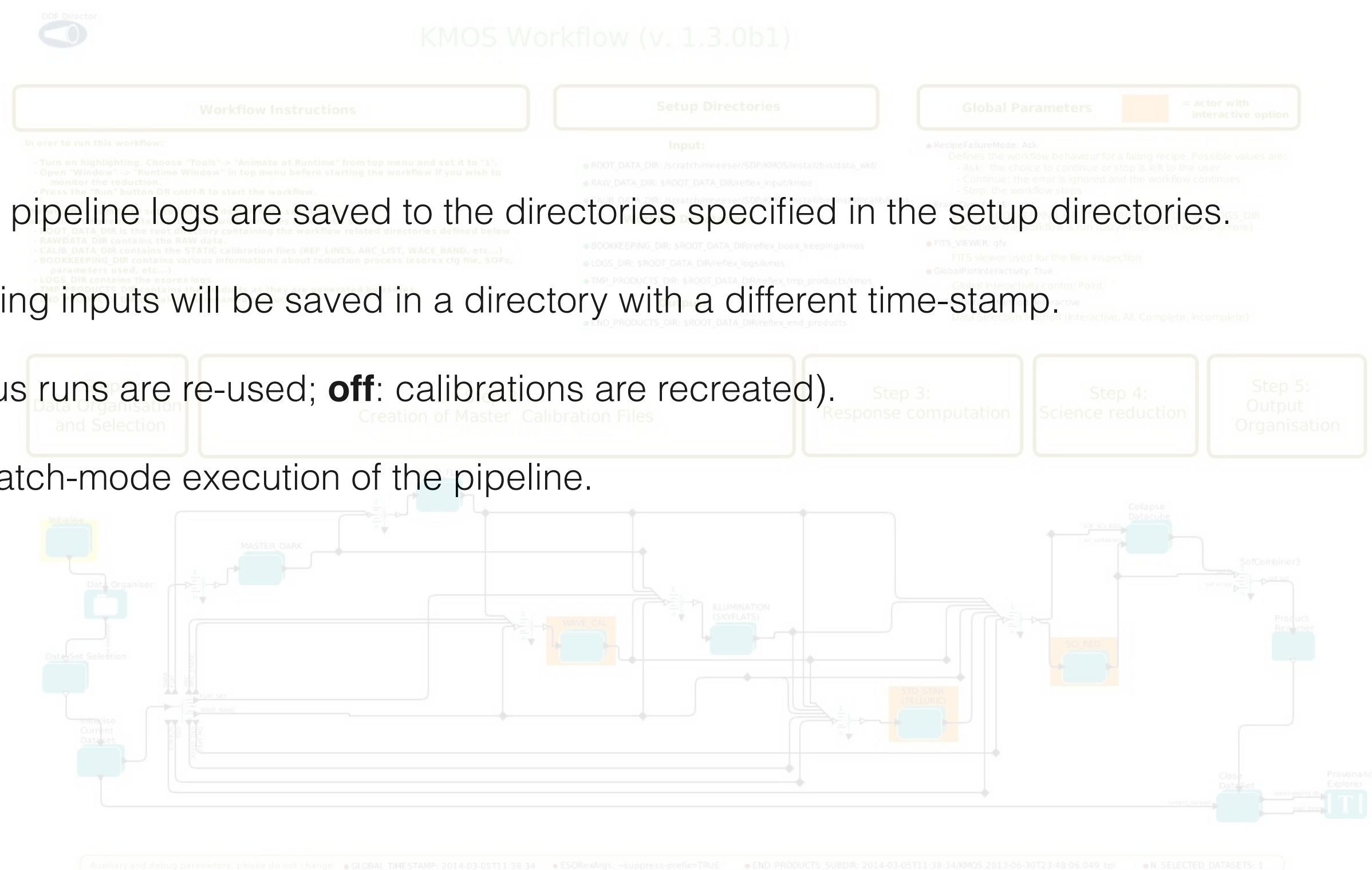
What if you want to add your own procedures to an existing Reflex workflow?

- it is possible to add your own scripts to the Reflex workflow . . . additional pipeline recipes, your own C routine, IDL script, or, easiest of all a Python routine.

- e.g. execute generic python code:

  - translate from/to Reflex ports to/from python script arguments a special syntax is used, with the help of an exisiting python module:

    In an open Reflex workflow, select from the components menu:
    ESO-reflex —> Scripting.kar —> PythonActor
    (I will give a demonstration of this in the next session)

For more details see:

Reflex Workflow Development Guide: ftp://ftp.eso.org/pub/dfs/reflex/reflex_dev_guide-1.0.pdf
Reflex User Manual: ftp://ftp.eso.org/pub/dfs/reflex/ReflexUserManual-3.5.pdf
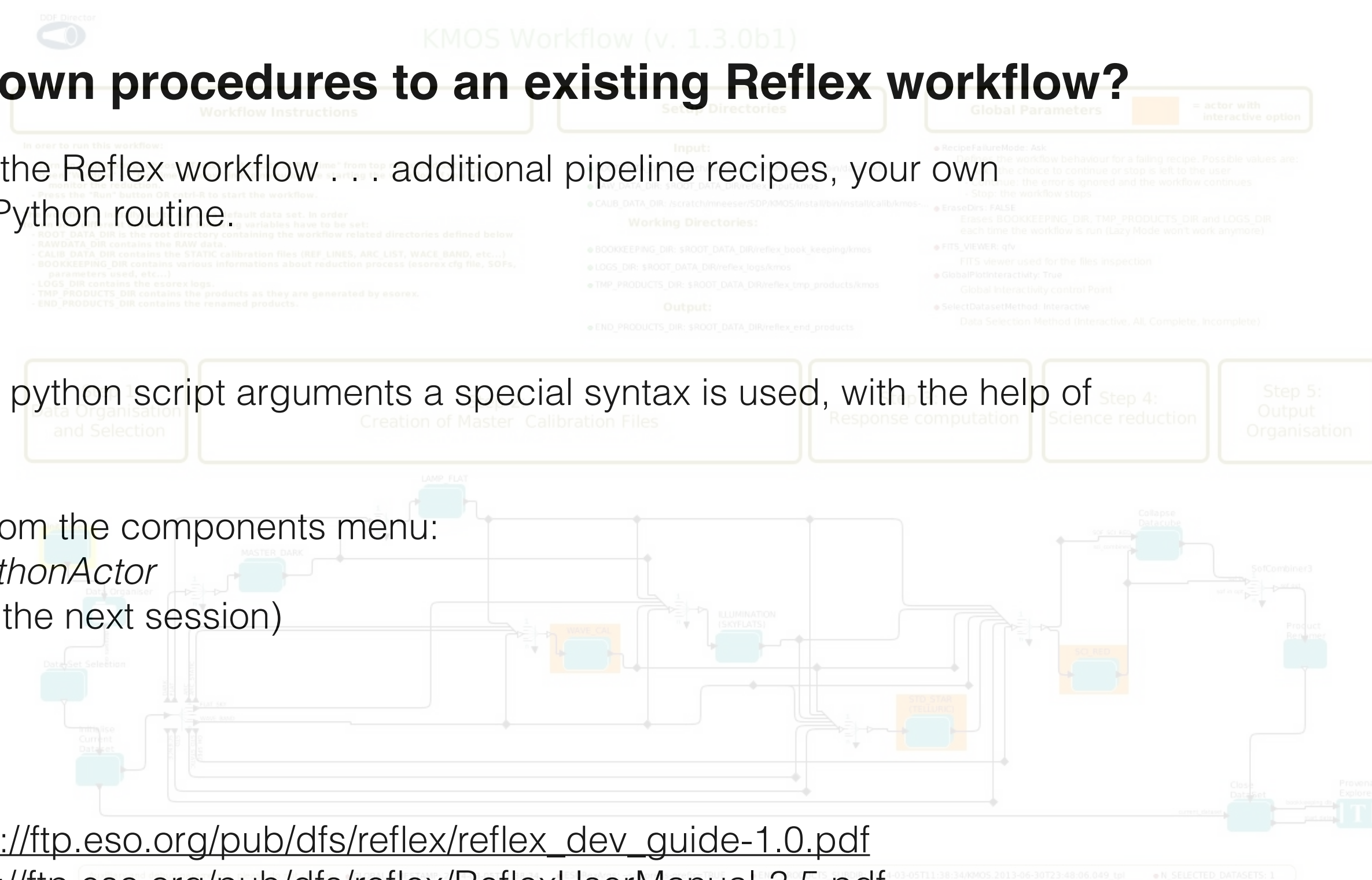
The ESO Archive and Reflex Pipeline Workflows

(or, how do I get data and process it too?)

Mark Neeser (ESO)

Outline

Part 1: In Theory

Getting Data from ESO:
- the contents of the ESO Science Archive Facility (SAF)
- how to access the archive for high-level science data or for raw data

Using Data from ESO:
- Reflex workflows and ESO pipelines
- Reflex workflow basics
- installing Reflex
- a summary of the main features of Reflex

Part 2: In Practise

A Practical Demonstration of:
- retrieving raw data from the ESO archive
- retrieving science products from the ESO archive
- running the HAWK-I Reflex workflow

A Demonstration of the ESO Archive

A Demonstration of the HAWK-I Reflex Workflow